UNIVERSITY OF TORONTO
3 1761 01185144 1

Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

http://www.archive.org/details/electronmatters00richuoft

CAMBRIDGE PHYSICAL SERIES

THE
ELECTRON THEORY
OF
MATTER

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
London: FETTER LANE, E.C.
Edinburgh: 100 PRINCES STREET

London: H. K. LEWIS, 136 GOWER STREET, W.C.
London: WILLIAM WESLEY AND SON, 28 ESSEX STREET, STRAND
Berlin: A. ASHER AND CO.
Leipzig: F. A. BROCKHAUS
New York: G. P. PUTNAM'S SONS
Bombay and Calcutta: MACMILLAN AND CO., LTD.
Toronto: J. M. DENT AND SONS, LTD.
Tokyo: THE MARUZEN-KABUSHIKI-KAISHA

All rights reserved

THE
ELECTRON THEORY
OF
MATTER

BY

O. W. RICHARDSON
M.A. (CAMB.), D.SC. (LOND.), F.R.S.
WHEATSTONE PROFESSOR OF PHYSICS AT KING'S COLLEGE, LONDON

Cambridge:
at the University Press
1914

LIBRARY
29
MAY
1978
UNIVERSITY OF TORONTO

Cambridge:
PRINTED BY JOHN CLAY, M.A.
AT THE UNIVERSITY PRESS

QC
793
.5
E62R5

PREFACE

THIS book is based on a course of lectures which I have delivered to graduate students at Princeton University during the last few years. My aim has been to exhibit the extent to which the fundamental facts of physical science may be coordinated by means of the conception of the electron and the laws of electrodynamics. In developing the subject I have started from the most elementary beginnings, and I have therefore found it necessary to include much matter which is to be found in any ordinary text-book of the theory of electricity and magnetism. It is hoped that the lack of conciseness thereby involved may be more than atoned for by the wider circle to which the book may appeal. The course of lectures at Princeton on which the book is founded proved useful as an introduction to the methods of modern mathematical physics in addition to forming a presentation of the results of recent physical discovery.

The broad scope of the subject makes it imperative that a good deal of selection should be exercised as to the nature and treatment of the topics considered. In determining these, consideration has been given to importance, interest, and instructiveness, roughly in the order named. The necessary incompleteness is remedied to some extent by references to scientific papers and to other works. These references are intended to supplement the discussion in the text rather than to exhibit the historical development of the subject. Thus many important papers are not referred to. I have tried, however, to be as accurate as possible in any statements which deal specifically with historical matters.

For a variety of reasons the book has, unfortunately, suffered considerable delay in passing through the press. I have, however, found it possible to incorporate some account of the important recent results while correcting the proofs; so that, with some reservations, the book may be regarded as fairly representing the state of the subject up to the time of printing. The most serious exception is in the part dealing with the electron theory of metallic conduction, where the important theoretical papers of Keesom (*Communications from the Leiden Physical Laboratory*, Supplement Nos. 30 and 32 (1913)) and Wien (*Columbia University Lectures* (1913) and *Berlin Sitzungsber.* p. 184 (1913)) as well as the full account of Kamerlingh Onnes' experimental work at low temperatures (*Comm. Leiden*, Supplement No. 34 (1913) and No. 139 (1914)) did not reach me in time to be dealt with. These papers lead one to hope that the difficulties which beset the electron theory of metallic conduction in its usual form may be overcome by the application of the ideas underlying Planck's theory of radiation. In any event the theories of Chapters XVII and XVIII should be valid at sufficiently high temperatures when the results of the quantum theory coalesce with those of the continuous theory. Many other branches of the subject are in a similar, though possibly less aggravated, situation; amongst these the questions of atomic structure, spectroscopic emission, X-rays and the magnetic properties of bodies are conspicuous examples. At the present time this field is unquestionably a very fruitful one both for the experimental and for the theoretical physicist.

I am indebted to the kindness of the publishers of the following journals and works for permission to reproduce various diagrams, viz. *Annalen der Physik*, Figs. 25, 30 *a*, 32 and 42; *Journal de Physique*, Figs. 35, 48 and 49; *Philosophical Magazine*, Figs. 52 and 56; Lorentz' *Theory of Electrons*, Fig. 55; Stark's *Principien der Atomdynamik*, Figs. 53 and 54.

I wish to express my thanks to Dr C. J. Davisson of the Carnegie Institute, Pittsburg, and to Dr K. T. Compton of the Reed College, Oregon, for the assistance I have received from the notes they took of my lectures, as well as to Mr T. G. Bedford, the Editor of the Cambridge Physical Series, for his valuable help and suggestions in reading the proofs.

O. W. RICHARDSON.

KING'S COLLEGE, LONDON.
May, 1914.

CONTENTS

CHAP.		PAGE
I.	The origin of the electron theory	1
II.	Electric intensity and potential	12
III.	Dielectric media	39
IV.	The electron theory of dielectric media	58
V.	Magnetism	77
VI.	Electromagnetism	86
VII.	Electromagnetic waves	115
VIII.	Dispersion, absorption and selective reflexion	142
IX.	The fundamental equations	182
X.	The activity of the forces	201
XI.	Charged system in uniform motion	217
XII.	Charge moving with variable velocity	243
XIII.	The aether	268
XIV.	The principle of relativity	296
XV.	Radiation and temperature	326
XVI.	The theory of magnetism	361
XVII.	The kinetic theory of electronic conduction	398
XVIII.	The equilibrium theory of electronic conductors	441
XIX.	Types of radiation	476
XX.	Spectroscopic phenomena	512
XXI.	The structure of the atom	549
XXII.	Gravitation	588
	Index of names	606
	Index of subjects	609

CHAPTER I

THE ORIGIN OF THE ELECTRON THEORY

The Electron Theory of Matter may be looked upon as a form of atomic theory. It differs from the form of the atomic theory with which chemistry is familiar, especially in that it makes the ultimate atoms minute geometrical configurations of electric charge, instead of particles of uncharged matter. A large number of different lines of inquiry, often closely interwoven, have led to the adoption of such a view of the structure of matter. Of these different lines of inquiry, however, three may be considered pre-eminently conspicuous.

In the first place, although the electron theory has made most rapid progress in the last two decades, it is a logical development of the views held a century ago by Davy* and Berzelius† and especially of the views to which Faraday‡ was led by his electrochemical discoveries made somewhat later. Davy concluded, from a general review of the electrochemical phenomena known in his day, that the forces between the chemical atoms were of electrical origin. Shortly afterwards a complete system of chemical structure depending on the same idea was developed by Berzelius and, although in its original form Berzelius's electrochemical theory was insufficiently elastic, its main features have much in common with the most modern views on the subject. The laws of electrolysis discovered by Faraday led to an important advance by pointing distinctly to an atomic constitution for electricity; for they showed that each chemical atom invariably transported either a definite quantity of electricity or an integral multiple of that quantity, a multiple which was determined by the chemical valency of the atom. This inference from Faraday's electrolytic researches was strongly advocated much later by von Helmholtz*.

---

* *Phil. Trans.* p. 1 (1807).

† *Mem. Acad. Stockholm* (1812); *Nicholson's Journal*, vols. xxxiv. and xxxv. (1813).

‡ *Exp. Res.* §§ 377, 523, 661, 713, 821 and especially 852, 869.

The second line of inquiry referred to dealt mainly with optical phenomena. It would be impossible adequately to discuss at this stage the complex questions which present themselves in this connection. It may be permissible to recall that Maxwell's electromagnetic theory had been found to account satisfactorily for the behaviour of light, and electromagnetic waves in general, in free space; but that difficulties presented themselves when phenomena like refraction and dispersion, which depend upon transmission through material media, were considered. These were found capable of removal by introducing the simple hypothesis that the material media contained particles having appropriate natural frequencies of vibration. Maxwell's electromagnetic theory of light naturally suggested that these particles were electrically charged, and the facts of dielectric polarisation were then found to fall into line, approximately at any rate, with the optical phenomena. The theory of the propagation of light in moving media also made important advances under the influence of the new views. A striking confirmation of the correctness of the general position was furnished by Zeeman's discovery of the change of the frequency of spectral lines when the emitting source was placed in a strong magnetic field. The magnitude of this change enabled an estimate of the ratio of the charge to the mass of the particles to be made. The resulting values were in substantial agreement with those which were obtained at about the same time by entirely different and more direct methods. In the development of the theory along the lines just indicated the ideas of H. A. Lorentz† and J. Larmor‡ have had a preponderating influence.

The third line of attack was furnished by the experimental study of the phenomena accompanying the discharge of electricity through gases and especially of the properties of the cathode rays, the Roentgen rays and the rays emitted by radioactive substances. The matter described in the ensuing paragraph will serve to give a general idea of the kind of information which has been supplied by these researches.

---

* "Faraday Lecture" (1881).

† *Arch. Neerl.* vol. xxv. p. 363 (1892). *Theory of Electrons*, Leipzig, 1909.

‡ *Phil. Trans.* A. vol. CLXXXV. p. 821 (1894). *Aether and Matter*, Cambridge, 1900.

The Isolation of the Elementary Charge.

The electron theory may now be said to have developed far beyond the region of hypothesis. Discovery after discovery during the last fifteen years has established indubitably the existence of a negative electron whose properties are independent of the matter from which it originates. J. J. Thomson[*] and Wiechert[†], independently, showed that the magnetic deflection of the cathode rays present in a vacuum tube at a low pressure proved that they consisted of negatively charged particles for which the value of e/m, where e is the electric charge and m the mass of one of the particles, was equal to about $1{\cdot}8 \times 10^7$ E.M. units. About the same time Lorentz[‡] showed that Zeeman's[§] discovery of the shift of the spectral lines of an emitting gas, produced by a magnetic field, pointed to the existence, within the atom, of negatively charged particles which had approximately the same value of e/m. Now the value of e/m for the lightest known chemical atom, the atom of hydrogen, can be obtained very accurately by electrolytic experiments and is found to be equal to $9{\cdot}577 \times 10^3$ E.M. units. Hence it would follow that if the charge carried by a cathode ray particle or by the particles which emit the spectral lines were identical with the charge carried by a hydrogen atom in electrolysis, as, indeed, we should rather expect would be the case from Faraday's electrolytic experiments, then the mass of these particles must be very much less than that of a hydrogen atom.

This question was soon put to the test of experiment. C. T. R. Wilson[‖] had shown that when moist ionised gas is subjected to sudden expansion, a cloud forms and the drops of water condense on the ions in preference to the uncharged molecules. J. J. Thomson[¶] utilised this phenomenon in order to count the number of ions in a volume of gas containing a measured total charge. Stated briefly the method is as follows: The amount of water condensed on the whole of the drops in the cloud may be calculated from the degree of supersaturation produced by the known expansion. The application of Stokes's formula for the rate of fall of a sphere in a viscous fluid gives the average size of each drop. Thus these data determine the total number n of the drops. It is assumed, for sufficient reasons, that practically all the drops contain one and only one ion. The total charge ne on all the ions could be determined by sweeping them out of the chamber into an electrometer before the expansion took place, the strength of the source of ionisation being the same as in the condensation experiments. Thus the charge on a single ion was obtained by division. Proceeding in this way Thomson showed that the negative ions liberated in air by Roentgen rays and by the β rays from radium each carried the same charge as the hydrogen ion in electrolysis.

---

[*] *Phil. Mag.* V. vol. XLIV. p. 298 (1897).

[†] *Verhandl. der Physik.-ökon. Gesellsch. zu Königsberg. i. Pr.* (1897).

[‡] *Phil. Mag.* V. vol. XLIII. p. 232 (1897).

[§] *Zittingsversl. der Akad. van Wet. te Amsterdam*, vol. v. pp. 181, 242 (1896).

[‖] *Roy. Soc. Proc.* March 19, 1896.

[¶] *Phil. Mag.* V. vol. XLVI. p. 528 (1898).

The ions investigated in these experiments are rather complicated structures and are not identical with the electron. In the case of the ionisation produced by ultra-violet light falling on a metal it was shown by Thomson that the particles when first emitted have the same value of e/m as the cathode rays. These would not be likely to aggregate together in the presence of gas molecules, and C. T. R. Wilson* showed that the negative ions from ultra-violet light behaved exactly like those from the other ionising agents in his condensation experiments. The inference from these experiments therefore is that the particles which form the cathode rays and which are emitted during photoelectric action carry a charge equal to that of the hydrogen atom in electrolysis. Experiments by Townsend†, on the rate of fall of the clouds produced when the gases evolved from chemical actions occurring in the wet way are allowed to bubble through water, had previously led him to conclude that the ions present in such gases carry the same charge as a hydrogen ion in electrolysis.

This conclusion has been strengthened by other methods of determining the charge on an electron. One of these depends on the theory of the radiation of electromagnetic energy from hot bodies. The theory of this method will be considered in the sequel[*]. One of the recent methods, which is due to Rutherford[†] and Geiger, depends upon the properties of radioactive substances. These are found to emit positively charged bodies, called α particles, which carry twice the charge e of an electron and are able to produce a large number of new ions when they pass through a gas. By magnifying this secondary ionisation by means of an auxiliary electric field and also using a very sensitive electrometer, Rutherford was able to detect the ionisation produced by a single α particle. When a very weak radioactive preparation was used the α particles were emitted at times separated by rather wide and irregular intervals, and as the effect produced by each one separately could be detected, the number emitted by a given amount of the radioactive substance in a given time could be measured. The only other datum which is required to measure e is the quantity of positive electricity which is carried away from the same quantity of the preparation by the α rays. This had previously been obtained by other experiments.

---

* *Phil. Trans.* A. vol. CXCII. p. 403 (1899).

† *Phil. Mag.* Feb. 1898.

The falling drop method has recently been improved by H. A. Wilson[‡] and R. A. Millikan[§]. The former showed that the charge on the drops could be deduced from the rate of fall under gravitation combined with different electric fields, without making use of the degree of supersaturation; whilst the latter showed how the drops of water could be replaced by drops of a non-volatile oil. The drops of oil have the great advantage that they do not evaporate: and by allowing a sufficient number of electrons to combine with them and applying a supporting electric field which just balances the gravitational force, they can be kept under observation for an indefinite length of time. In this way Millikan has shown that the method is capable of yielding results of very great precision.

All the three methods last mentioned are quite accurate and exhibit an excellent agreement. It is claimed that the charge e on an electron is known to within 1 per cent. Millikan's[‖] latest value is $e = 4{\cdot}81 \times 10^{-10}$ E.S. unit or $1{\cdot}60 \times 10^{-20}$ E.M. unit.

---

[*] See chap. xv.

[†] *Roy. Soc. Proc.* A. vol. LXXXI. pp. 141, 163 (1908).

[‡] *Phil. Mag.* VI. vol. v. p. 429 (1903).

[§] *Phil. Mag.* VI. vol. XIX. p. 209 (1910).

[‖] *Phys. Rev.* vol. XXXIV. p. 399 (1912).

Application to the Atomic Theory.

These experiments have led to other important consequences. Since the charge carried by one gram atom of an element is accurately known from electrolytic experiments, and the charge carried by a monovalent electrolytic ion has been shown to be equal to that of a gaseous ion, it follows that the number of atoms in one gram atom of any substance is known to the same degree of accuracy as e. Since the charge which is carried by one gram atom of a monovalent element in electrolysis is $9{\cdot}649 \times 10^3$ E.M. units it follows that the number of atoms in one gram atom of any element is $6{\cdot}02 \times 10^{23}$. Also, since the charge required to liberate half a cubic centimetre of H_2 at 0° C. and 760 mms. is 0·4327 E.M. unit, it follows that the number of molecules in one cubic centimetre of any gas under standard conditions of temperature and pressure is

$$\frac{0{\cdot}4327}{1{\cdot}60} \times 10^{20} = 2{\cdot}70 \times 10^{19}.$$

These values are in agreement with the comparatively inaccurate estimates which had previously been given by methods based on the kinetic theory of gases and other considerations.

Millikan was also able to observe the changes produced by the combination of single ions with the drops. These experiments, as well as those of Rutherford with the α rays, furnish a very direct and convincing proof of the atomic theory of matter and electricity. The consequences of the atomic theory of matter have recently been strikingly verified by experiments in other directions. Perrin* has shown that the irregular motions of minute suspended particles in fluids are in accordance with the requirements of the kinetic theory of gases. A study of these motions also leads to a determination of the number of molecules in one gram molecule of any element and thus to a determination of e. The value obtained by Perrin is in agreement with the other recent determinations. An examination of the distribution of velocity and kinetic energy among the electrons emitted by hot bodies, which has been carried out by the writer, partly in collaboration with F. C. Brown†, has shown that the motions of these electrons are in very close accordance with those required by Maxwell's theory, for the molecules of a gas of equal molecular weight.

---

* *Annales de Chim. et de Phys.* 1909.

† *Phil. Mag.* 1908 and 1909.

Electromagnetic Inertia.

About thirty years ago J. J. Thomson* pointed out the extremely important result that an electric charge possessed inertia simply in virtue of the energy of its electromagnetic field, and he succeeded in calculating the magnitude of the electrical inertia, or electromagnetic mass, as it is usually called, of a charged sphere. This additional mass does not play any important part in ordinary electrostatic experiments, as it is always small compared with the mass of the uncharged portion of the conductors and insulators which are experimented with. The case is very different when we are dealing with charged particles whose mass is only one eighteen-hundredth of that of an atom of hydrogen. It was obviously an important experimental problem to determine how much of the inertia of the electron was of the type foreseen by Thomson and how much, if any, was to be attributed to "ordinary" mechanical mass. Fortunately the two kinds of mass differentiate themselves rather clearly. The mechanical mass is supposed to be independent of the velocity of the body, following the principles of mechanics laid down by Newton, whereas the electromagnetic mass is a continuous function of the velocity and approaches infinity as the velocity of the electric charge approaches that of light.

The experimental problem was resolved by Kaufmann†, who measured the value of e/m for the electrons emitted by radium bromide, some of which have velocities as high as $2{\cdot}89 \times 10^{10}$ centimetres per second. He showed that the mass of these electrons varied with the speed, and in fact in a manner very similar to that predicted by Thomson. His final conclusion was that if there was any part of the mass of an electron which was ordinary mechanical mass it was very small in comparison with the part which was of electromagnetic origin.

The Negative Electron.

We have now succeeded in the isolation of a charged particle whose mass is much less than that of any known chemical atom. So far as our experiments enable us to tell, the whole of this mass arises from the electric charge the particle carries. There can be no question but that this is the negative electron. The structure of the positive electricity which goes to make up the remainder of the uncharged chemical atom is still uncertain; but the results of experiments so far point to both the charge and mass of the positive electrons being different from those of the negative. There is no evidence, so far as the writer is able to observe, which supports the view that the positive electron is, as it were, a reflection of the negative.

---

* *Phil. Mag.* V. vol. xi. p. 229 (1881).

† *Ann. der Phys.* vol. xix. p. 487 (1906); cf. also H. Starke, *Ver. der Deutsch. Physik. Ges.* vol. v. p. 241 (1903).

The foregoing considerations enable us to define more precisely the use of the word electron. In future we shall restrict the term to particles which consist of a geometrical configuration of electricity and nothing else, whose mass, that is, is all electromagnetic. For a particle which is a charged molecule or atom, that is to say, a molecule or atom which has lost or gained one or more negative electrons, we shall use the term ion. A wider meaning than this is currently attributed to the word ion, in the sense of any charged particle which is considered to have a separate existence or which behaves as a dynamical unit. Some of these are comparatively large bodies and contain very many atoms or molecules. To distinguish them from the smaller ions already referred to, Stark has suggested the use of the terms molion and atomion. As, however, we shall not have to consider the large molions we shall simply use the word ion instead of atomion.

According to the view we are developing, all interactions between material systems result from the electrical charges which make up their ultimate parts. The space in the neighbourhood of an electric charge is to be looked upon as having properties different from that some distance away, since an electric charge of the same sign is repelled with a greater force in the former case than in the latter. This state of things is described by saying that the electric charge is surrounded by a field of force.

It is often convenient to attribute this field of force to disturbances produced by the electric charge in a medium, the aether, which fills all space. Looked at in this way the real electron, the part which acts, is the surrounding aether which is outside its geometrical boundary; and the electron theory is the science of the properties of the aether, of which the electric charges are local modifications.

Different Elements of Electricity.

Our ignorance of the geometrical distribution of the electrification constituting an electron is almost complete, but this is not a serious disadvantage in considering many applications of the theory. It will often be sufficient to regard an electron as a point charge having a definite inertia coefficient or mass. In such cases the mode of distribution of the electric charge, whether it is a point, line, surface or volume charge and whether it is distributed with spherical or linear symmetry or not, is unimportant, provided that it is confined to a minute region of space. Although this is often true there are some investigations for which the ultimate geometrical distribution of the electrification is important; as for instance in the case already mentioned of the calculation of the electromagnetic mass. These two distinct classes of cases require as a rule quite different methods of attack.

These remarks will make it clear why it is necessary to have different elementary portions of electricity in different investigations dealing with the electron theory. In the first place we may have to consider the forces acting on or arising from a small portion of the electron itself. The elementary quantity of electricity concerned here may be denoted by $\rho\, d\tau$, where ρ is the volume density of the electricity at the point of the electron under consideration and $d\tau$ is an element of volume of the latter, infinitesimal in comparison with the size of the electron. In other investigations our element of electricity will be the charge on a single electron whose value e is determined by the equation

$$e = \iiint \rho\, d\tau,$$

where the volume integral extends over the volume occupied by the electron. In still another class of investigations the volume element of electric charge will occupy a region of space containing an enormous number of electrons both negative and positive. In this case also it is conveniently represented by $\rho\, d\tau$, or when confusion is likely to occur with the first case by $\bar{\rho}\, d\tau$. If n is the number of negative electrons per unit volume at any point and e is the charge carried by each, N and E being the corresponding quantities for the positive electrons, then $\bar{\rho}\, d\tau = (NE + ne)\, d\tau$.

This is the kind of volume element which occurs in the usual problems of electrostatics, where effects arising from the discrete structure of the electrification and from the electronic structure of matter are smoothed out. It will be observed that each of the three elements of electricity is of a successively higher order of magnitude, both as regards distribution in space and in reference to the quantity of electricity it contains.

Force between Electrically Charged Masses.

On the electron theory the interaction between material bodies carrying electric charges is to be pictured in a manner somewhat different from that usual in electrostatics. The forces act on each single electron present in the two bodies, irrespective of whether it may be regarded as forming part of the free electric charge, or whether it is simply one of the constituent electrons in the matter which carries the electric charge. The field of force arising from a single electron is assumed to obey the same laws as that arising from a small charged particle in the theory of electrostatics.

Let us consider the forces acting between two charged material particles situated at the points P and Q at a distance r apart. Suppose that the matter at P consists of N_1 positive electrons of charge E_1 and n_1 negative electrons of charge e_1, the matter at Q being composed of N_2 positive electrons of charge E_2, and n_2 negative electrons of charge e_2. The force exerted by P on Q or vice versa will consist of the algebraic sum of the repulsions between all the like electrons and the attractions between all the unlike electrons. It will thus be made up of the four items which follow :

1. The repulsion of the positive electrons. Any one electron at Q exerts a force E_1E_2/r^2 on each individual positive electron at P. The total force arising in this way from each electron at Q and acting on P will therefore be $N_1E_1E_2/r^2$; hence the force due to all the electrons at Q is $N_1N_2\dfrac{E_1E_2}{r^2}$.

2. The repulsion of the negative electrons. This clearly amounts to $n_1n_2\dfrac{e_1e_2}{r^2}$.

3. The attraction of the positive electrons at P for the negative electrons at Q. This is readily seen by similar reasoning to amount to $N_1n_2\dfrac{E_1e_2}{r^2}$.

4. The attraction of the positive electrons at Q for the negative electrons at P. This is evidently equal to $N_2 n_1 \dfrac{E_2 e_1}{r^2}$. Thus the total repulsion between P and Q amounts to

$$N_1 N_2 \frac{E_1 E_2}{r^2} + n_1 n_2 \frac{e_1 e_2}{r^2} + N_1 n_2 \frac{E_1 e_2}{r^2} + N_2 n_1 \frac{E_2 e_1}{r^2}$$
$$= (N_1 E_1 + n_1 e_1)(N_2 E_2 + n_2 e_2)/r^2.$$

If q_1, q_2 represent the magnitude and sign of the free electricity at P and Q respectively, then

$$q_1 = N_1 E_1 + n_1 e_1 \text{ and } q_2 = N_2 E_2 + n_2 e_2,$$

so that the repulsive force is equal to $q_1 q_2 / r^2$, the usual electrostatic law.

This result may easily be extended to the general case where there are different kinds of electrons of the same sign provided with different charges.

If $N_1 E_1 = n_1 e_1$ and $N_2 E_2 = n_2 e_2$, the bodies are uncharged and the force between them vanishes. Thus the above formulation does not leave any room for the explanation of gravitational attraction between uncharged material particles. This lacuna is considered in Chapter XXII.

CHAPTER II

ELECTRIC INTENSITY AND POTENTIAL

That branch of electrical science which deals with the properties of electrical charges when at rest is called electrostatics. It is the oldest branch of electricity, some of the fundamental phenomena of frictional electricity having been known qualitatively by the ancients. By charges "at rest" we mean at rest relatively to one another. We shall see that there is no evidence for the view that the absolute motion of the charges affects their action on one another. When the charges move relatively to one another, important differences are observed which will be considered later.

From the point of view of electrostatics the most interesting properties of electrified bodies are their mutual repulsions and attractions. It is found that all electrified bodies can be grouped into two classes such that all the bodies of either class repel all the other bodies in the same class but attract all the bodies in the other class. A body is said to be positively or negatively electrified according to the class to which it belongs. The distinction is not merely one of sign, since, as we shall see in the sequel, there are important qualitative differences between positive and negative electricity.

In the previous chapter we have stated that the force between two charged bodies of sufficiently minute size is proportional to the product of their charges divided by the square of the distance between them. This law of variation of force with distance was discovered by Priestley in 1767 and rediscovered by Coulomb in 1785*. We may express it in the form $F = k\frac{ee'}{r^2}$, where k is a constant quantity whose magnitude depends on our definition of the unit of electric charge. In dealing with electrical phenomena we shall assume that k is a universal constant independent of the sign or other quality of e or e', although we shall see that if we are to account for gravitation on the electron theory this cannot be the case. The omission will, however, make no practical difference in the case of purely electrical effects: it is only when gravitational effects are concerned that the difference is important.

---

* Cf. Whittaker, *History of Theories of the Aether and Electricity*, Dublin, 1910, pp. 50, 56.

The magnitude of our unit of electric charge will depend on how we determine the constant k which is of unknown dimensions since the dimensions of e are unknown. The dimensions of Fr^2 are of course perfectly determinate and equal to ML^3T^{-2}. We shall fix our units by the convention that if the charges are separated only by ordinary space, then F is in dynes and r in centimetres provided $k = 1/4\pi$. This is equivalent to defining our unit of electric charge as that which repels an equal and similar charge at unit distance with a force of $1/4\pi$ dynes. This unit of charge was introduced by Heaviside and differs from the ordinary electrostatic unit which makes $k = 1$ and which we made use of in the last chapter. The new unit has certain advantages in improving the symmetry of formulae which we shall obtain later.

As the dimensions of k are unknown, it is sometimes inadvisable to suppress it in our formulae even if we have given it a definite numerical value. This is particularly true in investigations of a very fundamental character. In such cases it is convenient to write $F = \frac{1}{4\pi}\frac{ee'}{Kr^2}$, where K has the value $\frac{1}{4\pi}$ on the ordinary electrostatic system of units, the value unity on Heaviside's electrostatic system and may take other values on other systems of units. K is sometimes called the dielectric coefficient or specific inductive capacity of the aether.

In dealing with electrostatics it is not necessary for us to determine how it comes about that two electric charges attract or repel one another. Two entirely different attitudes towards this and the cognate question in regard to gravitational attraction have been adopted by different schools of thought. One school, adopting the dogma of "action at a distance," holds that the law of force between charges is the fundamental thing and that it is useless to attempt to go beyond it. The other denies the possibility of action at a distance and derives the law of force between charges from the effect of the charged bodies on an intervening medium, the aether. There is no *a priori* reason for adopting one view rather than the other, although most of the great investigators have been ranged against the action at a distance school. The great advantage of the medium view is that it pictures the operation of a mechanism whose consequences insist on foretelling themselves, whereas the other is mere dead description. For this reason the medium view has been most successful in leading the great advances in electrical science, whilst the strength of the action at a distance formulation lies in its mathematical simplicity. There is no absolute contradiction between the two views; Maxwell has shown that a system of possible stresses in the medium will give rise to the observed attractions and repulsions. Which is the more desirable is largely a matter of taste or convenience. This is particularly the case so far as electrostatic phenomena are concerned. When we come to the consideration of electromagnetic phenomena we shall see that the medium view possesses important advantages, in certain directions, at any rate.

Electric Intensity.

We shall now suppose that the space surrounding an electric charge is different from that elsewhere. We do not need to consider how this is brought about. It may be that the charge produces a change in the state of the surrounding aether; or the charge may have parts which extend into the region about it; or it may be merely a manifestation of a hyperspatial mechanism; or it may even be something which is incapable of description in mechanical terms. The important point is that if another charge is placed at any point of such space it will be acted on by a force and accelerated. The force acting on this second charge is proportional jointly to its magnitude e and to a vector E determined by the first charge. This is true provided that there are no other charges in the field. In that case E will be compounded of the effects due to the various charges other than e. The vector E is called the electric intensity at the point of the field under consideration.

The function E is what is known as a vector point function; in other words it is a function which for each point of space has both magnitude and direction. A function which has only one magnitude at each point may be called a single-valued point function. The electric intensity E is single-valued in this sense; but its direction is indefinite at the points where E vanishes.

Clearly the electric intensity at a point distant r from a point charge e' is in Heaviside's units $E = e'/4\pi r^2$. This follows from the inverse square law of force. The electric intensity due to a complicated distribution of charge may obviously be obtained by the integration of the amount arising from each volume element. In making the calculation it is necessary to integrate for each component of the electric intensity separately and combine the results according to the rule for the composition of forces. This resolution and subsequent composition of vectors is often troublesome as well as clumsy, and it is not necessary for the calculation of the electric intensity. It may be dispensed with by the introduction of another function known as the Potential.

The Potential.

The Potential is defined as the work divided by the charge when an infinitesimal electric charge is brought from some standard position to the point in question. The standard position is usually taken to be a point at an infinite distance away from charged bodies. The value of the potential calculated in this way must be independent of the path of approach to the point under consideration, otherwise an indefinite amount of work could be obtained by making the charge move round a closed contour passing through the point under consideration and the standard position. This would be contrary to the law of conservation of energy. The potential is a function of each point in space and possesses magnitude but not direction. Such a function is known as a *scalar* point function. The electrostatic potential is single-valued.

Let P, Q be two points at an infinitesimal distance ds apart and such that PQ is in the direction of the resultant electric intensity E at P. Let dV be the increase in the potential in passing from P to Q and let the direction $P \rightarrow Q$ be considered positive. Then

$dV = -E ds$, so that $E = -\frac{dV}{ds}$. The value of the component of electric intensity at P in any other direction whose inclination to PQ is θ will be $E_\theta = E \cos\theta$. This will always be less than E. If ds' denotes an infinitesimal length laid off in the direction making the angle θ with PQ, we shall have $E_\theta = -\frac{\partial V}{\partial s'}$. It is clear from this that the direction of the resultant electric intensity is the direction of quickest diminution of V. Starting out from P let us lay off a length PQ in the direction of the resultant intensity at P, then from Q a length QR in the direction of the resultant intensity at Q and so on from point to point. In this way we shall draw a curved line in space such that the tangent to the curve at any point will give the direction of the resultant electric intensity at that point. The curve will also represent the path of a positive charge devoid of inertia which moves under the influence of the field. Such a line is called a line of electric force. If the direction cosines of the resultant electric intensity at any point are l, m, n, we shall have $E_x = lE$, $E_y = mE$, $E_z = nE$, where E_x, E_y, E_z are the components of E. If $ds = (dx, dy, dz)$ is the element of arc of the line of force at the same point,

$$l = \frac{dx}{ds}, \quad m = \frac{dy}{ds}, \quad n = \frac{dz}{ds},$$

so that

$$\frac{dx}{E_x} = \frac{dy}{E_y} = \frac{dz}{E_z} = \frac{ds}{E} \quad \text{.....................}(1).$$

These are the differential equations of a line of force.

If an electric charge moves always at right angles to the lines of force, no work will be done on it, so that all the points on the surface on which it moves will be at the same potential. A surface traced out in this way is called an equipotential surface. It is clear that the equipotential surfaces always cut the lines of force normally and that the whole field may be divided up into a series of right prismatic cells by means of a system of equipotential surfaces intersecting orthogonally a series of tubular surfaces containing the lines of force. The latter surfaces are called tubes of force.

In general we may write the value of the potential at any point P as $V = -\int E \cos\theta \, ds$, where E is the resultant electric intensity at any point of the path (ds) of integration and θ is the angle that E makes with ds. The limits of integration are from infinity to the point P. If l, m, n are the direction cosines of ds we have $E\cos\theta = lE_x + mE_y + nE_z$, so that

$$V = -\int (lE_x + mE_y + nE_z)\,ds = -\int (E_x dx + E_y dy + E_z dz),$$

taken between the same limits. For a point charge e the electric intensity at a point distant r is radial and equal to $e/4\pi r^2$; so that the potential V is $e/4\pi r$, as is clear upon integrating along a radial path from infinity to the point P.

The following considerations enable us to find the value of the potential V due to a number of point charges e_0, e_1, e_2, etc. Consider the intensity at $Q(x, y, z)$ arising from one of the charges e_0 at $P(a_0, b_0, c_0)$. The resultant intensity is $\dfrac{e_0}{4\pi r_0^2}$, where $r_0^2 = (x-a_0)^2 + (y-b_0)^2 + (z-c_0)^2$. The x component, X_0, of this is

$$\frac{e_0}{4\pi r_0^2}\,\frac{x-a_0}{r_0} = -\frac{\partial}{\partial x}\left(\frac{e_0}{4\pi r_0}\right).$$

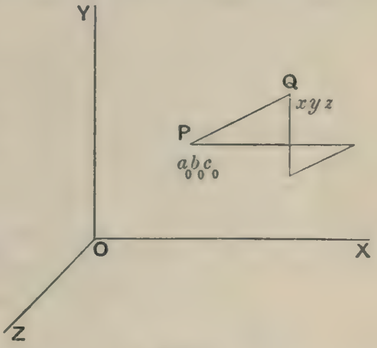


Fig. 1.

Similarly the x component of the intensity which arises from any of the other charges e_n distant r_n from x, y, z may be written

$$X_n = \frac{e_n}{4\pi r_n^2}\,\frac{x-a_n}{r_n} = -\frac{\partial}{\partial x}\left(\frac{e_n}{4\pi r_n}\right).$$

But the x component X of the electric intensity due to all the charges is

$$X = X_0 + X_1 + X_2 + \ldots = -\frac{\partial}{\partial x}\left(\frac{e_0}{4\pi r_0} + \frac{e_1}{4\pi r_1} + \frac{e_2}{4\pi r_2} + \ldots\right) = -\frac{\partial V}{\partial x}.$$

So that the potential V due to a number of point charges is $\dfrac{1}{4\pi}\Sigma\dfrac{e}{r}$, and is equal to the sum of the potentials due to the separate charges.

We may now deduce the differential equation which is satisfied by the potential. Consider again the potential V_0 at Q due to the charge e_0 at P. Since

$$\frac{\partial}{\partial x}\left(\frac{1}{r_0}\right) = -\frac{1}{r_0^2}\frac{\partial r_0}{\partial x} = -\frac{1}{r_0^2}\frac{x-a_0}{r_0}$$

we have
$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r_0}\right) = -\frac{1}{r_0^3} + 3\frac{(x-a_0)^2}{r_0^5}.$$

Similarly
$$\frac{\partial^2}{\partial y^2}\left(\frac{1}{r_0}\right) = -\frac{1}{r_0^3} + 3\frac{(y-b_0)^2}{r_0^5}$$

and
$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{r_0}\right) = -\frac{1}{r_0^3} + 3\frac{(z-c_0)^2}{r_0^5}.$$

So that
$$\frac{\partial^2}{\partial x^2}\left(\frac{e_0}{4\pi r_0}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{e_0}{4\pi r_0}\right) + \frac{\partial^2}{\partial z^2}\left(\frac{e_0}{4\pi r_0}\right) = 0.$$

If we have charges e_1, e_2, etc. at points distant r_1, r_2, etc. from Q, similar equations will hold for each of them, and, by addition, we have
$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\left(\frac{e_0}{4\pi r_0} + \frac{e_1}{4\pi r_1} + \frac{e_2}{4\pi r_2} + \dots\right) = 0.$$

But the potential V at the point Q is that due to all the charges e_0, e_1, e_2, etc. It is therefore equal to $\frac{e_0}{4\pi r_0} + \frac{e_1}{4\pi r_1} + \frac{e_2}{4\pi r_2} + \dots$. We thus find that the potential V must satisfy the equation
$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 \quad\text{.....................}(2).$$

This is often written $\nabla^2 V = 0$ and sometimes $\Delta V = 0$. This result will clearly hold so long as Q lies outside all of the charges e_0, e_1, etc. We shall see that when Q lies inside a charged body the differential equation is modified. The equation $\nabla^2 V = 0$ is known as Laplace's equation, and the operator $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ as Laplace's operator.

The result $V = \frac{1}{4\pi}\Sigma\frac{e}{r}$ can readily be extended from the case of a series of point charges to that of a continuous distribution of electricity. For the space occupied by the latter can be split up into an indefinitely large number of volume elements $d\tau$. The charge on each of these is $\rho\, d\tau$, if ρ is the volume density of the electrification. The potential due to the distribution, at a point outside it, will clearly be equal to $\frac{1}{4\pi}\iiint\frac{\rho}{r}\,d\tau$, where the triple integral extends over the whole of the electrified body and r is the distance of any element $\rho\, d\tau$ from the external point. If there are also surface distributions of electricity these will evidently also add to the potential an amount $\frac{1}{4\pi}\iint \frac{\sigma}{r}\,dS$, where σ is the surface density of the charge at the element dS and r its distance from the external point.

The Potential at internal points.

Further consideration is necessary before we can apply the foregoing expression to the potential at points inside the electrified medium. At such points the denominator r becomes zero, and we must be certain that this does not make the integral which represents the potential infinite. The matter may be investigated as follows. About the internal point Q describe a sphere of infinitesimal radius ϵ. We shall suppose the density ρ of the charge to be everywhere finite and the radius ϵ to be chosen so small that the density of the charge varies continuously throughout the volume of the sphere. This condition can always be satisfied. The potential at Q will consist of two parts: (1) V_1, arising from the charge outside the sphere, and (2) V_2, arising from the charge inside the sphere. The former is clearly finite. Let ρ_m be the maximum value of ρ inside the sphere. Then $V_2 \ngtr \frac{1}{4\pi}\iiint_\epsilon \frac{\rho_m}{r}\,d\tau$. The element of volume $d\tau$ in polar coordinates is $r^2 dr \sin\theta\, d\theta\, d\phi$. So that

$$V_2 \ngtr \frac{\rho_m}{4\pi}\int_0^\epsilon r\,dr\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi} d\phi \ngtr \tfrac{1}{2}\rho_m\epsilon^2.$$

This vanishes when ϵ is made sufficiently small; so that we conclude that the charge in the immediate neighbourhood contributes nothing to the potential at any point. The formula $V = \frac{1}{4\pi}\iiint \frac{\rho}{r}\,d\tau$ therefore holds generally both for points outside and for points inside the electrified medium.

The Derivatives of V.

The electric intensity $-\frac{\partial V}{\partial r}$ and all higher derivatives of V contain r to a higher order in the denominator than does V itself. They are therefore all finite at external points.

The electric intensity is also finite at internal points. Consider again the small sphere of radius ϵ and divide V into two parts

V_1 and V_2 as before. As we have just shown, $\dfrac{\partial V_1}{\partial r}$ is finite. We have

$$\frac{\partial V_2}{\partial r} = \frac{1}{4\pi}\iiint_\epsilon \left\{\frac{1}{r}\frac{\partial \rho}{\partial r} - \frac{\rho}{r^2}\right\} d\tau.$$

By taking ϵ small enough ρ/r^2 can be made as large as we please compared with $\dfrac{1}{r}\dfrac{\partial \rho}{\partial r}$, so that the latter may be disregarded in the limit. Thus

$$-\frac{\partial V_2}{\partial r} \ngtr \frac{\rho_m}{4\pi}\int_0^\epsilon dr \int_0^\pi \sin\theta\, d\theta \int_0^{2\pi} d\phi \ngtr \rho_m \epsilon,$$

which vanishes when $\epsilon = 0$. We conclude therefore that the electric intensity at internal points may be obtained by differentiating the potential.

Gauss's Theorem.

The consideration of the distribution of the electric intensity in relation to surfaces in space leads to interesting results. Consider any surface whatever and let dS be an element of it. Let N be the component of the electric intensity at dS, along the normal to the element, N being reckoned positive if it is in the direction of the normal drawn outward from the surface. We shall now prove that the integral $\iint N dS$ taken over any closed surface is equal to e, where e is the algebraic sum of the charges enclosed within the surface.

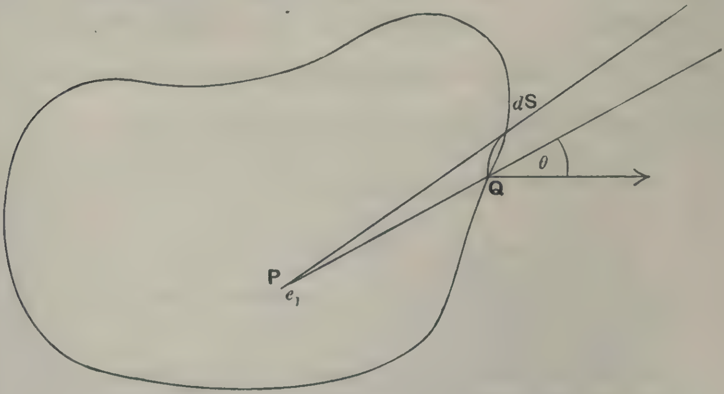


Fig. 2.

Consider first the intensity E_1 at Q due to a charge e_1 at a point P within the surface. We have $E_1 = e_1/4\pi PQ^2$ and

$$N_1 dS = \frac{e_1}{4\pi PQ^2} dS \cos\theta,$$

where θ is the angle between PQ and the normal to dS. Now $\dfrac{dS\cos\theta}{PQ^2}$ is the solid angle $d\omega_1$ subtended by dS at P. Thus, so far as the single charge e_1 is concerned, we have $N_1 dS = \dfrac{e_1}{4\pi} d\omega_1$. If there are a large number of point charges e_1, e_2, etc., the resultant normal component of electric intensity

$$N = N_1 + N_2 + N_3 + \dots,$$

so that $$4\pi N\, dS = e_1 d\omega_1 + e_2 d\omega_2 + e_3 d\omega_3 + \dots.$$

Now $$\iint e_1 d\omega_1 = 4\pi e_1, \quad \iint e_2 d\omega_2 = 4\pi e_2, \text{ etc.}$$

So that $$\iint N dS = e_1 + e_2 + e_3 + \dots = e \quad \dots\dots\dots\dots\dots(3).$$

This result can obviously be extended from a series of point charges to a continuous distribution in the same manner as that employed in dealing with the potential.

It remains to prove that charges outside the closed surface contribute nothing to the surface integral. It is evident that every conical element of solid angle $d\omega$ arising from an external charge will cut the closed surface an even number of times. The value of $N dS$ for the intersections of the cone by the surface will be alternately positive and negative since the direction of the electric intensity is constant in space but alternates in sign with reference to the successive normals. The numerical value of $N dS$ is the same for successive intersections, being equal to $\dfrac{e}{4\pi} d\omega$. So that the surface integral is divided up into a series of pairs of equal and opposite elements. Its value is therefore zero, and we conclude that the value of $\iint N dS$ over any closed surface is equal to the charge inside.

This result is known as Gauss's Theorem.

The theorem is of great value as a means of calculating the value of the electric intensity arising from various symmetrical distributions of electric charge. Thus in the case of a uniformly charged spherical shell the intensity at any point external to the shell must be the same at every point on the spherical surface through the point and concentric with the shell. It also follows from the symmetry of the problem that the intensity must be radial. Since the area of a sphere of radius r is $4\pi r^2$, the electric intensity E at a distance r from the centre of the charged shell is given by $4\pi r^2 E = e$; so that

$$E = e/4\pi r^2 \qquad \text{.........................(4).}$$

In a similar way we may prove that the force vanishes inside the shell. These results may be extended to the case of a sphere charged throughout its volume so that layers equidistant from the centre are charged to equal density. Thus we may show, for example, that the force *inside* a solid uniform sphere of electricity varies as the distance from the centre.

The application of Gauss's Theorem to the tubes of force mentioned on p. 16, is instructive. As we have seen, a tube of force is a tubular region bounded by a surface which is the envelope of the lines of force. Let us apply Gauss's Theorem to a portion of such a tube, terminated at each end by equipotential surfaces. The lines of force run along the tubular surfaces so that at each point the component of the intensity normal to these surfaces vanishes. Over the ends the resultant electric intensity will be normal to the surfaces. Let it be E_1 at the end where the cross section is S_1, E_2 and S_2 being the corresponding quantities at the other end. The value of $\iint N dS$ over the whole surface considered is clearly $E_1 S_1 - E_2 S_2$. If the tube of force is in a region where there are no electric charges this vanishes, so that $E_1 S_1 = E_2 S_2$; thus the electric intensity at any point is inversely as the area of cross section of the tubes of force at that point.

Under the conditions contemplated in electrostatics the surface of a conductor of electricity must be an equipotential surface; otherwise there would be currents of electricity flowing from one part of the surface to another. The tubes of force must therefore start normally from such a surface. Now apply Gauss's Theorem to the region bounded by a tube of force and its continuation into the substance of the conductor and terminated by equipotential surfaces, one inside and the other outside the conductor. The electric intensity vanishes over all the surface inside the conductor and the normal component vanishes over the tubular surface outside. The value of $\iint N dS$ is thus equal to the value $E_1 S_1$ of this quantity over the end section. This is equal to the charge inside; which is σS_2, where σ is the surface density of the charge on the conductor and S_2 the area of its intersection by the tube of force. We therefore have $E_1 S_1 = \sigma S_2$. If we make the end section approach indefinitely near to the charged surface, $S_1 = S_2$, so that

$$E = \sigma \qquad \text{.........................} \quad (5).$$

Thus the electric intensity at a point just outside a charged conductor is normal to the surface and equal to the charge per unit area of it. This result is known as Coulomb's Law.

We may express Gauss's Theorem analytically as a relation between a surface and a volume integral. If ρ is the volume density of the electrification at any point inside a closed surface the total charge inside the surface will be $\iiint \rho \, dx dy dz$. The normal component of the electric intensity outside the surface is $N = -\frac{\partial V}{\partial n}$. Gauss's Theorem may therefore be expressed in the form

$$\iiint \rho \, dx dy dz = -\iint \frac{\partial V}{\partial n} dS \qquad \text{................} (6).$$

In a region where there are no charges, if l, m, n are the direction cosines of the normal to any closed surface, we have

$$\iint (lE_x + mE_y + nE_z)\, dS = \iint N dS = 0.$$

A vector E whose components E_x, E_y, E_z satisfy the relation $\iint (lE_x + mE_y + nE_z)\, dS = 0$ over any closed surface is said to be solenoidal. Thus the electric intensity in free aether is a solenoidal vector. We shall see that there are other solenoidal vectors in the theory of electricity and magnetism whose properties are of great importance.

Green's Theorem.

An important theorem discovered by George Green* in 1828 enables us to express a volume integral taken throughout an enclosed space in terms of surface integrals over the boundaries of the space. This theorem, which is named after the discoverer, is a purely geometrical theorem, but it has many important applications in the theory of electricity.

---

* "An essay on the application of Mathematical Analysis to the theories of Electricity and Magnetism," by George Green, published at Nottingham in 1828.

Consider any closed space bounded by any number of closed surfaces, which may be either internal or external. As a particular case the external boundary may be a sphere of infinite radius. Let u, v, w be any continuous point function. Then Green's Theorem states that

$$\iiint \left\{\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right\} dx\,dy\,dz = + \iint (lu + mv + nw)\, dS,$$

where the volume integral is taken throughout the space between the boundaries and the surface integral over the bounding surfaces, and l, m and n are the direction cosines of the normals to the surfaces drawn away from the space throughout which the volume integral is taken.

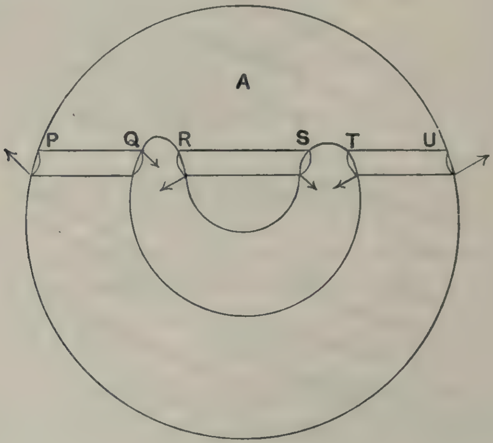


Fig. 3.

Let the two parallel lines $PQRSTU$ represent the section, by the plane of the paper, of a right prism whose section by the plane yOz would be $dydz$. Consider the contribution to the integral $\iiint \frac{\partial u}{\partial x}\, dx\,dy\,dz$ throughout the space A arising from this prismatic portion of it. This will clearly be

$$dy\,dz\,(-u_P + u_Q - u_R + u_S - u_T + u_U),$$

where u_P, u_Q, etc. are the values of u at P, Q, etc. Now

$$dy\,dz = -l_P dS_P = +l_Q dS_Q = -l_R dS_R = +l_S dS_S = -l_T dS_T = +l_U dS_U,$$

where $l_P dS_P$ are the values of l and dS at P, and so on. The alternation of signs is due to the fact that the normals alternately make acute and obtuse angles with the positive direction of x. So that

$$\iiint \frac{\partial u}{\partial x}\, dx\, dy\, dz = +\iint lu\, dS \quad\text{..................(7)}.$$

Similar equations hold for v and w, so that, by addition,

$$\iiint \left\{\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right\} dx\, dy\, dz = +\iint \{lu + mv + nw\}\, dS \quad\text{...(8)}.$$

Now let $u = U\dfrac{\partial V}{\partial x}$, $v = U\dfrac{\partial V}{\partial y}$, $w = U\dfrac{\partial V}{\partial z}$.

After substitution we have

$$\begin{aligned}\iiint U\left\{\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right\} dx\, dy\, dz \\ + \iiint \left\{\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z}\right\} dx\, dy\, dz \\ = +\iint U\left\{l\frac{\partial V}{\partial x} + m\frac{\partial V}{\partial y} + n\frac{\partial V}{\partial z}\right\} dS \\ = +\iint U\frac{\partial V}{\partial n}\, dS \quad\text{..............................(9)},\end{aligned}$$

where $\dfrac{\partial V}{\partial n}$ denotes the rate of change of V along the normal at dS.

If we substitute $V\dfrac{\partial U}{\partial x}$ for u, $V\dfrac{\partial U}{\partial y}$ for v, and $V\dfrac{\partial U}{\partial z}$ for w, we obtain similar expressions in which U and V are interchanged. Subtracting the two equations we have

$$\begin{aligned}\iiint \{U\nabla^2 V - V\nabla^2 U\}\, dx\, dy\, dz \\ = +\iint \left\{U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n}\right\} dS \quad\text{.........(10)}.\end{aligned}$$

The three equations preceding, (8), (9) and (10), are all different forms of Green's Theorem.

An important case arises when we put $U =$ constant in equation (9) or (10). We then have

$$\iiint \left\{\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right\} dx\, dy\, dz = +\iint \frac{\partial V}{\partial n}\, dS \quad\text{......(11)}.$$

We may apply this result to a minute cavity of any shape inside a charged body.

We have seen (p. 23) that

$$\iint \frac{\partial V}{\partial n_e} dS = -\iiint \rho \, dx \, dy \, dz.$$

Hence $$\iiint \left\{\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right\} dx \, dy \, dz = -\iiint \rho \, dx \, dy \, dz \quad \ldots(12).$$

Since this equality is independent of the shape and size of the cavity it follows that

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\rho \quad \ldots\ldots\ldots\ldots\ldots\ldots(13).$$

This equation is known as Poisson's equation. It is the general form of the differential equation satisfied by the potential. At points where there are no electric charges we have $\rho = 0$, so that we get Laplace's equation $\nabla^2 V = 0$ as a particular case of it.

E_x, E_y, E_z, the components of E, are given by

$$E_x = -\frac{\partial V}{\partial x}, \quad E_y = -\frac{\partial V}{\partial y}, \quad E_z = -\frac{\partial V}{\partial z}.$$

This is often written $E = (E_x, E_y, E_z) = -\text{grad } V$. We may write the equation $\nabla^2 V = -\rho$ or 0 in the form

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho.$$

This equation is often abbreviated to $\text{div } E = \rho$, the operator div acting upon any vector denoting the sum of the results of the action of the operator grad on each of the corresponding components of the vector.

Transformation of Laplace's Operator.

In dealing with certain problems, particularly in cases possessing spherical or cylindrical symmetry, it is not desirable to use Poisson's and Laplace's equations in rectangular coordinates. Polar or cylindrical coordinates are much more suitable. The transformation of the equations to these new coordinates can very easily be effected by making use of the theorem that the surface integral of $\frac{\partial V}{\partial n_e}$ over any closed surface is equal to the volume integral of the volume density inside (Gauss's Theorem, p. 23). To illustrate the method we shall first apply it to the simpler case of rectangular coordinates.

Consider the element of volume $dx\,dy\,dz$, in rectangular coordinates, whose centre is x, y, z. The coordinates of the angular points of the element are $x-\frac{1}{2}dx$, $y-\frac{1}{2}dy$, $z-\frac{1}{2}dz$, etc. Let V be the value of the potential at x, y, z. Then the mean value of the potential over the right hand face will be $V+\frac{1}{2}\frac{\partial V}{\partial x}dx$ and that over the left hand face will be $V-\frac{1}{2}\frac{\partial V}{\partial x}dx$. The mean value of the component of the electric intensity along the outward drawn normal at the right hand face $=-\frac{\partial}{\partial x}\left(V+\frac{1}{2}\frac{\partial V}{\partial x}dx\right)$ and at the left hand face $=+\frac{\partial}{\partial x}\left(V-\frac{1}{2}\frac{\partial V}{\partial x}dx\right)$. The flux over these faces, as we may call, for the sake of brevity, the surface integral of the normal intensity, is $dy\,dz\times-\left[\frac{\partial V}{\partial x}+\frac{1}{2}\frac{\partial^2 V}{\partial x^2}dx\right]$ in the one case and $dy\,dz\times\left[\frac{\partial V}{\partial x}-\frac{1}{2}\frac{\partial^2 V}{\partial x^2}dx\right]$ in the other. The total flux over the pair of faces perpendicular to Ox is therefore

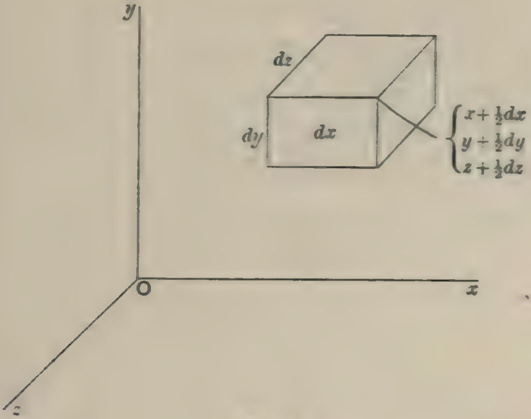


Fig. 4.

$$-\frac{\partial^2 V}{\partial x^2}dx\,dy\,dz.$$

Similar expressions hold for the Oy and Oz faces; so that the total flux for the whole cube is

$$-\left\{\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}\right\}dx\,dy\,dz.$$

By Gauss's Theorem this is equal to the charge inside or $\rho\, dx\, dy\, dz$. We thus derive $\nabla^2 V = -\rho = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}$, which is Poisson's Equation in rectangular coordinates.

Polar Coordinates.

Let r, θ, ϕ be the polar coordinates of the centre of the element of volume under consideration, ϕ being the azimuth measured from a fixed plane passing through the polar axis and θ the angle the radius makes with the polar axis Oz.

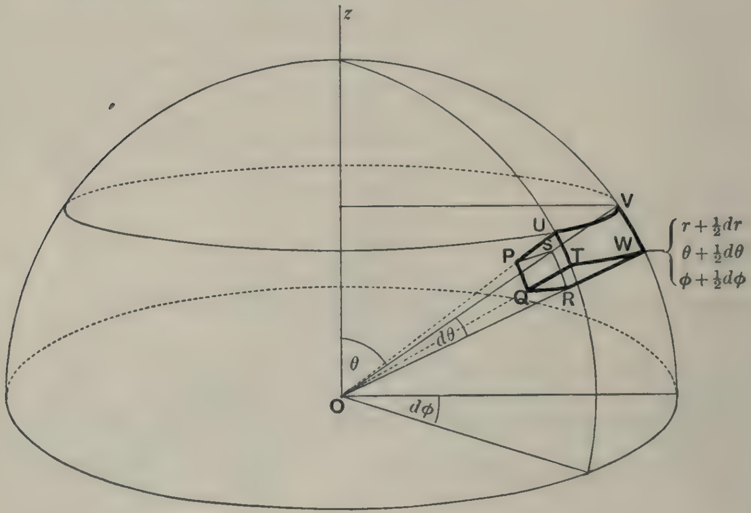


Fig. 5.

Let $PQRSVWTU$ be the element of volume. It is formed by the intersection of the following surfaces: (1) two spheres centre O and radii $r - \frac{1}{2}\, dr$ and $r + \frac{1}{2}\, dr$; (2) two planes passing through Oz making angles $\phi - \frac{1}{2}\, d\phi$ and $\phi + \frac{1}{2}\, d\phi$ with a fixed plane passing through the same axis; (3) two cones described coaxally about the axis Oz and of semi-angles $\theta - \frac{1}{2}\, d\theta$ and $\theta + \frac{1}{2}\, d\theta$ respectively. The coordinates of the angular points of the elements of volume are $P = r - \frac{1}{2}\, dr$, $\theta - \frac{1}{2}\, d\theta$, $\phi - \frac{1}{2}\, d\phi$, and so on. The radius of the circle formed by the intersection of the cone of semi-angle θ and the sphere of radius r is clearly $r \sin\theta$. The volume of the element is

$$dr \times r\, d\theta \times r \sin\theta\, d\phi = r^2 dr \sin\theta\, d\theta\, d\phi.$$

Consider first the flux over the faces which are perpendicular to r. The potential at the centre being V, the mean potential over the outer face will be $V + \frac{1}{2}\frac{\partial V}{\partial r}dr$ and over the inner face $V - \frac{1}{2}\frac{\partial V}{\partial r}dr$, the normal intensities being $-\frac{\partial}{\partial r}\left(V + \frac{1}{2}\frac{\partial V}{\partial r}dr\right)$ and $+\frac{\partial}{\partial r}\left(V - \frac{1}{2}\frac{\partial V}{\partial r}dr\right)$. The areas of these faces are respectively $(r \pm \frac{1}{2}dr)\sin\theta\, d\phi \times (r \pm \frac{1}{2}dr)\, d\theta$ or $r(r \pm dr)\sin\theta\, d\theta\, d\phi$ to the first order. The fluxes are therefore

$$\left(-\frac{\partial V}{\partial r} - \frac{1}{2}\frac{\partial^2 V}{\partial r^2}dr\right) \times (r + dr)\, r\sin\theta\, d\theta\, d\phi$$

and
$$\left(+\frac{\partial V}{\partial r} - \frac{1}{2}\frac{\partial^2 V}{\partial r^2}dr\right) \times (r - dr)\, r\sin\theta\, d\theta\, d\phi.$$

The total flux over these two faces is thus

$$-r^2 dr\sin\theta\, d\theta\, d\phi\left\{\frac{\partial^2 V}{\partial r^2} + \frac{2}{r}\frac{\partial V}{\partial r}\right\}.$$

The mean potentials at the conical surfaces $PSVU$ and $QRWT$ are $V - \frac{1}{2}\frac{\partial V}{\partial \theta}d\theta$ and $V + \frac{1}{2}\frac{\partial V}{\partial \theta}d\theta$ respectively. The corresponding normal intensities are obtained by differentiation with respect to the element of arc $r\, d\theta$, and are therefore equal to $+\frac{1}{r}\frac{\partial}{\partial \theta}\left(V - \frac{1}{2}\frac{\partial V}{\partial \theta}d\theta\right)$ and $-\frac{1}{r}\frac{\partial}{\partial \theta}\left(V + \frac{1}{2}\frac{\partial V}{\partial \theta}d\theta\right)$. The area of $PSVU$ is

$$dr \times r\sin(\theta - \tfrac{1}{2}d\theta)\, d\phi = r\,dr\,d\phi \times [\sin\theta\cos\tfrac{1}{2}d\theta - \cos\theta\sin\tfrac{1}{2}d\theta]$$
$$= r\,dr\,d\phi\,(\sin\theta - \tfrac{1}{2}\cos\theta\, d\theta)$$

to the first order. Similarly,

$$QRWT = r\,dr\,d\phi\,[\sin\theta + \tfrac{1}{2}\cos\theta\, d\theta].$$

The fluxes therefore are

$$dr\,d\phi\left(\frac{\partial V}{\partial \theta} - \frac{1}{2}\frac{\partial^2 V}{\partial \theta^2}d\theta\right)(\sin\theta - \tfrac{1}{2}\cos\theta\, d\theta)$$

and
$$-dr\,d\phi\left(\frac{\partial V}{\partial \theta} + \frac{1}{2}\frac{\partial^2 V}{\partial \theta^2}d\theta\right)(\sin\theta + \tfrac{1}{2}\cos\theta\, d\theta).$$

The total flux is therefore

$$-r^2\,dr\sin\theta\,d\theta\,d\phi\left\{\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\cot\theta}{r^2}\frac{\partial V}{\partial\theta}\right\}.$$

The mean potentials of the plane surfaces *PQTU* and *RSVW* are $V+\frac{1}{2}\frac{\partial V}{\partial\phi}d\phi$ and $V-\frac{1}{2}\frac{\partial V}{\partial\phi}d\phi$ respectively. The element of arc normal to the planes being $r\sin\theta\,d\phi$, the mean normal intensities are

$$-\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\left(V+\frac{1}{2}\frac{\partial V}{\partial\phi}d\phi\right)\ \text{ and }\ +\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\left(V-\frac{1}{2}\frac{\partial V}{\partial\phi}d\phi\right).$$

The areas of the surfaces are equal to each other and to $dr\times r\,d\theta$. The fluxes are therefore equal to

$$-\frac{dr\,d\theta}{\sin\theta}\left(\frac{\partial V}{\partial\phi}+\frac{1}{2}\frac{\partial^2 V}{\partial\phi^2}d\phi\right)$$

and

$$+\frac{dr\,d\theta}{\sin\theta}\left(\frac{\partial V}{\partial\phi}-\frac{1}{2}\frac{\partial^2 V}{\partial\phi^2}d\phi\right),$$

their sum being

$$-r^2 dr\sin\theta\,d\theta\,d\phi\times\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}.$$

We thus find for the total flux over all the six surfaces of the element of volume

$$-r^2 dr\sin\theta\,d\theta\,d\phi\times\frac{\partial^2 V}{\partial r^2}+\frac{2}{r}\frac{\partial V}{\partial r}+\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\cot\theta}{r^2}\frac{\partial V}{\partial\theta}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}.$$

By Gauss's Theorem this is equal to the charge inside, which is

$$\rho\,d\tau=\rho\times r^2\sin\theta\,dr\,d\theta\,d\phi.$$

So that

$$\frac{\partial^2 V}{\partial r^2}+\frac{2}{r}\frac{\partial V}{\partial r}+\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\cot\theta}{r^2}\frac{\partial V}{\partial\theta}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}=-\rho\ldots(14).$$

This is therefore Poisson's Equation in spherical coordinates, and the operator

$$\frac{\partial^2}{\partial r^2}+\frac{2}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\cot\theta}{r^2}\frac{\partial}{\partial\theta}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\quad\ldots\ldots(15)$$

is the form which Laplace's operator ∇^2 takes in this system of coordinates.

Cylindrical Coordinates.

Any other system of orthogonal coordinates may be treated similarly. For instance in the case of cylindrical coordinates r, θ, z the element of volume is bounded by (1) two coaxal cylinders of radii r and $r+dr$ with their axes coincident with the axis of z, (2) two planes inclined at an angle $d\theta$ to one another and passing through the axis of z, (3) two parallel planes perpendicular to the axis of z and at a distance dz apart.

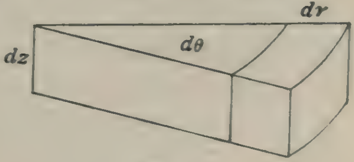


Fig 6.

The volume of the element is clearly $dr \times r d\theta \times dz$. The potential at the centre of the element being V, the flux over the outer cylindrical surface will be

$$\left(r+\frac{dr}{2}\right) d\theta\, dz \times -\frac{\partial}{\partial r}\left(V+\frac{1}{2}\frac{\partial V}{\partial r}dr\right),$$

and over the opposite face

$$\left(r-\frac{dr}{2}\right) d\theta dz \times \frac{\partial}{\partial r}\left(V-\frac{1}{2}\frac{\partial V}{\partial r}dr\right),$$

the total flux over the two faces being

$$-r dr\, d\theta dz \left\{\frac{\partial^2 V}{\partial r^2}+\frac{1}{r}\frac{\partial V}{\partial r}\right\}.$$

The area of the plane inclined faces is $dr\, dz$, the mean normal intensity over them $\mp\frac{1}{r}\frac{\partial}{\partial\theta}\left(V\pm\frac{\partial V}{\partial\theta}d\theta\right)$ and the total flux over them

$$-r\, dr\, d\theta\, dz \left\{\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}\right\}.$$

The area of the faces perpendicular to Oz is $dr \times r d\theta$ and the total flux over them $= -r\, dr\, d\theta\, dz\, \frac{\partial^2 V}{\partial z^2}$.

So that the flux over the whole six faces is

$$-r dr\, d\theta dz \left\{\frac{\partial^2 V}{\partial r^2}+\frac{1}{r}\frac{\partial V}{\partial r}+\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\partial^2 V}{\partial z^2}\right\} = \rho \times r dr\, d\theta dz.$$

Thus the form which Poisson's Equation takes in cylindrical coordinates is

$$\frac{\partial^2 V}{\partial r^2}+\frac{1}{r}\frac{\partial V}{\partial r}+\frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}+\frac{\partial^2 V}{\partial z^2} = -\rho \quad \text{............(16).}$$

The Uniqueness of the Solutions.

Laplace's Equation in spherical, cylindrical and rectangular coordinates is of the greatest importance in many branches of mathematical physics. It is clear from the preceding discussion that the first derivatives of the solutions of this equation represent a vector which flows out from a series of points uniformly in all directions. Its applicability to the theory of radiation, of conduction of heat and electricity, to hydrodynamics and gravitational attraction as well as to electrical and magnetic attractions is at once obvious.

The utility of the foregoing differential equations arises from the fact that if we can solve them we only need to be given the value of V over certain surfaces in order to obtain the complete distribution of electric force in the field. This result depends upon the theorem, which we shall now prove, that if V satisfies the equation $\nabla^2 V = -\rho$ throughout any region of space and has certain assigned values over surfaces bounding the region, then it is the only function which satisfies these conditions. For if not let V' also satisfy the same conditions and let us write $U = V \equiv V' - V$ in the expression for Green's Theorem in equation (9). Then ?

$$\iiint (V - V')\nabla^2(V - V')\,d\tau + \iiint\left[\left\{\frac{\partial(V - V')}{\partial x}\right\}^2 + \left\{\frac{\partial(V - V')}{\partial y}\right\}^2 + \left\{\frac{\partial(V - V')}{\partial z}\right\}^2\right]d\tau = \iint (V - V')\frac{\partial(V - V')}{\partial n}\,dS,$$

and since $\nabla^2 V' = \nabla^2 V = -\rho$ throughout the space and $V - V' = 0$ over the surfaces, we find

$$\iiint\left[\left\{\frac{\partial}{\partial x}(V - V')\right\}^2 + \left\{\frac{\partial}{\partial y}(V - V')\right\}^2 + \left\{\frac{\partial}{\partial z}(V - V')\right\}^2\right]dx\,dy\,dz = 0.$$

But this integral is a sum of squares; so it can only vanish if each term vanishes separately. We thus have

$$\frac{\partial}{\partial x}(V' - V) = \frac{\partial}{\partial y}(V' - V) = \frac{\partial}{\partial z}(V' - V) = 0.$$

Hence $V' - V =$ constant everywhere. But $V = V'$ on the surfaces; hence $V = V'$ everywhere.

Since $\iint (V - V') \frac{\partial}{\partial n} (V - V') \, dS$ also vanishes when $\frac{\partial V}{\partial n}$ is given over the surface S, it follows that V is unique except for an additive constant when the value of ρ is assigned throughout the space and that of $\frac{\partial V}{\partial n}$ over the boundaries; so that, in this case also, the electric intensity is determined uniquely.

It would lead us too far afield to consider the functions (Fourier's Series, Spherical and Zonal Harmonics and Bessel's Functions) which are the solutions of Laplace's Equation appropriate to particular problems. For the development of this interesting subject the reader may be referred to Byerly's *Fourier's Series and Spherical Harmonics.*

Total Energy of a System of Charges.

We may find the total energy of a system of charged bodies in terms of their charges and potentials as follows. Since the potential at any point of the field is equal to $\iiint \frac{\rho}{r} d\tau + \iint \frac{\sigma}{r} dS$ taken over all the charged bodies in the field, it will be reduced to $1/n$ of its value if *all* the charges are reduced in the ratio $n:1$. Let n be any very large number, and suppose that initially all the charges are at an infinite distance from one another. Bring up $1/n$ of each element of charge to its final position. If V is the final potential at any point the potential will change during this operation from 0 to $\frac{V}{n}$. The work done in bringing up the element $\frac{1}{n} \rho \, d\tau$ will lie between 0 and $\frac{1}{n^2} V\rho \, d\tau$. The work done in bringing up $1/n$ of all the charges will lie between 0 and $\frac{1}{n^2} \iiint V\rho \, d\tau$. Now bring up a second nth part of all the charges. This will raise the potential at any point from V/n to $2V/n$, and the work done in this second stage will lie between $\frac{1}{n^2} \iiint V\rho \, d\tau$ and $\frac{2}{n^2} \iiint V\rho \, d\tau$. If this process is continued the work done in the sth stage will lie between

$$\frac{s-1}{n^2} \iiint V\rho \, d\tau \text{ and } \frac{s}{n^2} \iiint V\rho \, d\tau.$$

The total work done in bringing up the whole of the charges from a state of infinite dissemination will lie between

$$\frac{1}{n^2}(0+1+2+\ldots+\overline{n-1})\iiint V\rho\,d\tau$$

and $$\frac{1}{n^2}(1+2+3+\ldots+n)\iiint V\rho\,d\tau.$$

This is equal to the total potential energy of the system, which therefore lies between

$$\frac{n(n-1)}{2n^2}\iiint V\rho\,d\tau \text{ and } \frac{n(n+1)}{2n^2}\iiint V\rho\,d\tau.$$

When n is increased indefinitely each of these values coincides with $\frac{1}{2}\iiint V\rho\,d\tau$. This is the part due to the volume charges. If there are surface charges we shall have to add $\frac{1}{2}\iint \sigma V\,dS$. The complete expression for the total energy of any system of charges is therefore

$$W=\frac{1}{2}\iiint V\rho\,d\tau+\frac{1}{2}\iint V\sigma\,dS \quad\text{............(17).}$$

The Energy in the Field.

In the preceding paragraph we have deduced an expression for the energy in terms of the charges and their potentials. On the view that electrical actions are transmitted through a medium, we should expect that the energy would reside in the medium. It is easy to obtain from the equation (17) an expression which admits of this interpretation. Since $\rho=-\nabla^2 V$ and $\sigma=-\frac{\partial V}{\partial n_e}$, we have

$$W=-\frac{1}{2}\iiint V\nabla^2 V\,d\tau-\frac{1}{2}\iint V\frac{\partial V}{\partial n_e}\,dS.$$

But by Green's Theorem, allowing for the reversal of sign arising from the fact that the normal is now drawn *into* the space considered, this is equal to

$$W=\frac{1}{2}\iiint\left\{\left(\frac{\partial V}{\partial x}\right)^2+\left(\frac{\partial V}{\partial y}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right\}d\tau \quad\text{......(18)}$$

$$=\frac{1}{2}\iiint E^2\,d\tau \quad\text{....................................(19).}$$

So that the energy of the system is the same as if each element of the field contained an amount of energy $\frac{1}{2}E^2$ per unit volume.

Stresses in the Field.

Maxwell showed that the forces acting on any system of charged bodies could be attributed to a system of stresses in the medium in which they are embedded. The necessary and sufficient condition for this to be the case is evidently that the resolved part, in any direction, of the resultant of all the forces acting on the parts of the system, arising from systems external to it, should be expressible in the form of an integral over any surface which surrounds and isolates the system. The alternative possibility would imply that part of the force was not transmitted across the boundary, through the action of the parts of the medium on one another, but arose from so-called action at a distance.

Consider any surface S surrounding and isolating the system of static charges e_1. Let X_1 be the x component of the resultant force acting on e_1 arising from all external electrical systems. Then if V is the potential and ρ the volume density at any point

$$X_1 = -\iiint \frac{\partial V}{\partial x} \rho \, d\tau = +\iiint \frac{\partial V}{\partial x} \nabla^2 V \, d\tau,$$

where the integrations are extended throughout the volume enclosed by S. The volume integrals will be capable of transformation into integrals over the boundary surface S* if we can write

$$\frac{\partial V}{\partial x}\left\{\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right\}$$

in the form
$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

We have
$$\frac{\partial V}{\partial x}\frac{\partial^2 V}{\partial x^2} = \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\partial V}{\partial x}\right)^2,$$

$$\frac{\partial V}{\partial x}\frac{\partial^2 V}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial V}{\partial x}\frac{\partial V}{\partial y}\right) - \frac{\partial V}{\partial y}\frac{\partial^2 V}{\partial x \partial y}$$
$$= \frac{\partial}{\partial y}\left(\frac{\partial V}{\partial x}\frac{\partial V}{\partial y}\right) - \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\partial V}{\partial y}\right)^2,$$

$$\frac{\partial V}{\partial x}\frac{\partial^2 V}{\partial z^2} = \frac{\partial}{\partial z}\left(\frac{\partial V}{\partial x}\frac{\partial V}{\partial z}\right) - \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\partial V}{\partial z}\right)^2.$$

---

* Cf. the proof of Green's Theorem, p. 24, Chap. II.

Thus the integrand will be in the form desired if we put

$$P = \frac{1}{2}\left\{\left(\frac{\partial V}{\partial x}\right)^2 - \left(\frac{\partial V}{\partial y}\right)^2 - \left(\frac{\partial V}{\partial z}\right)^2\right\} = p_{xx} \quad \text{......(20),}$$

$$Q = \frac{\partial V}{\partial x}\frac{\partial V}{\partial y} = p_{yx} = p_{xy} \quad \text{......(21),}$$

$$R = \frac{\partial V}{\partial x}\frac{\partial V}{\partial z} = p_{zx} = p_{xz} \quad \text{......(22).}$$

By considering the components of the resultant force parallel to the y and z axes we should arrive at similar surface integrals involving

$$p_{yy} = \frac{1}{2}\left\{\left(\frac{\partial V}{\partial y}\right)^2 - \left(\frac{\partial V}{\partial z}\right)^2 - \left(\frac{\partial V}{\partial x}\right)^2\right\} \quad \text{......(23),}$$

$$p_{zz} = \frac{1}{2}\left\{\left(\frac{\partial V}{\partial z}\right)^2 - \left(\frac{\partial V}{\partial x}\right)^2 - \left(\frac{\partial V}{\partial y}\right)^2\right\} \quad \text{......(24),}$$

$$p_{zy} = \frac{\partial V}{\partial y}\frac{\partial V}{\partial z} = p_{yz} \quad \text{......(25),}$$

p_{xy} and p_{xz}.

In the new notation in terms of the p's, which will be familiar to students of elasticity, we may write

$$+X_1 = \iiint \left(\frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{yx}}{\partial y} + \frac{\partial p_{zx}}{\partial z}\right) d\tau$$

$$= \iint (l p_{xx} + m p_{yx} + n p_{zx})\, dS \quad \text{......(26).}$$

The last integral is taken over the enclosing surface, and l, m, n are the direction cosines of dS, drawn away from the enclosed volume. Similarly

$$+Y_1 = \iint (l p_{xy} + m p_{yy} + n p_{zy})\, dS \quad \text{......(27)}$$

and

$$+Z_1 = \iint (l p_{xz} + m p_{yz} + n p_{zz})\, dS \quad \text{......(28).}$$

If we adopt the standpoint that the action of the electric charges on one another is transmitted by the intervening medium, then $p_{xx}, p_{yy}, p_{zz}, p_{yx}, p_{zx}, p_{zy}$ are the six components of the stress which transmits the action. From the point of view of action at a distance these quantities, on the other hand, have no physical significance.

In order to obtain a more definite picture of the physical nature of the supposed stresses let us consider the case in which dS is part of an equipotential surface, so that its normal is tangential to a line of force. Let the resultant electric intensity at dS be E, then

$$-\frac{\partial V}{\partial x}=lE, \quad -\frac{\partial V}{\partial y}=mE \text{ and } -\frac{\partial V}{\partial z}=nE,$$

and

$$\begin{aligned} p_{xx}&=\tfrac{1}{2}E^2(l^2-m^2-n^2), & p_{yz}&=E^2mn,\\ p_{yy}&=\tfrac{1}{2}E^2(m^2-n^2-l^2), & p_{zx}&=E^2nl,\\ p_{zz}&=\tfrac{1}{2}E^2(n^2-l^2-m^2), & p_{xy}&=E^2lm.\end{aligned}$$

The components of the force per unit area across dS are respectively

$$\left.\begin{aligned} lp_{xx}+mp_{yx}+np_{zx}&=\tfrac{1}{2}lE^2\\ lp_{xy}+mp_{yy}+np_{zy}&=\tfrac{1}{2}mE^2\\ lp_{xz}+mp_{yz}+np_{zz}&=\tfrac{1}{2}nE^2\end{aligned}\right\} \quad \text{...............(29)}.$$

Thus the resultant traction is normal to dS. It is therefore directed along the line of force and is a tension of amount $\frac{1}{2}E^2$ per unit area.

Next suppose that dS is at right angles to an equipotential surface. Its direction cosines l', m', n' will then satisfy the relation

$$l'\frac{\partial V}{\partial x}+m'\frac{\partial V}{\partial y}+n'\frac{\partial V}{\partial z}=0.$$

The x component of the stress across dS is

$$\begin{aligned} &l'p_{xx}+m'p_{yx}+n'p_{zx}\\ &=l'\frac{1}{2}\left\{\left(\frac{\partial V}{\partial x}\right)^2-\left(\frac{\partial V}{\partial y}\right)^2-\left(\frac{\partial V}{\partial z}\right)^2\right\}+m'\frac{\partial V}{\partial x}\frac{\partial V}{\partial y}+n'\frac{\partial V}{\partial x}\frac{\partial V}{\partial z}\\ &=-\frac{1}{2}l'\left\{\left(\frac{\partial V}{\partial x}\right)^2+\left(\frac{\partial V}{\partial y}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right\}\\ &=-\tfrac{1}{2}l'E^2,\end{aligned}$$

since

$$\left\{l'\left(\frac{\partial V}{\partial x}\right)+m'\left(\frac{\partial V}{\partial y}\right)+n'\left(\frac{\partial V}{\partial z}\right)\right\}\frac{\partial V}{\partial x}=0.$$

In a similar manner we may show that the y and z components are equal respectively to $-\frac{1}{2}m'E^2$ and $-\frac{1}{2}n'E^2$.

Thus when dS is at right angles to an equipotential surface the resultant stress is again normal to it. Since the expressions now have a negative sign in front of them, they will represent a pressure, not a tension. There is thus at every point in the field a tension equal to $\frac{1}{2}E^2$ along the lines of force and a pressure of equal amount in every direction at right angles to them.

It may be shown that these stresses will keep the aether in equilibrium. They are not the most general system of stresses which are equivalent to the electric force on the system of charged bodies e_1. For we may clearly add to them any distribution of stress whose resultant is equal to zero when integrated over any closed surface surrounding e_1. They are, however, the only system in which the stress at any point is determined solely by the electric intensity at that point*.

It is well to point out that this interpretation of the forces as a system of stresses in the medium has only been shown to be a possible, not a necessary one.

---

* Cf. Jeans, *Electricity and Magnetism*, p. 146, and Maxwell, *Electricity and Magnetism*, 3rd ed. vol. I. p. 158.

CHAPTER III

DIELECTRIC MEDIA

CAVENDISH, and Faraday independently, showed that when a condenser was charged so that the potential difference between the plates had a certain fixed value, the charge on the plates depended on the insulating medium between them. This proved that the forces between electric charges depended not only on the magnitude of the charges, but also on the nature of the material separating them. These experiments are often regarded as disproving the dogma of action at a distance. They are not capable, however, of establishing this inference; all that they prove with certainty is that electric charges act on a *material* medium in such a way as to make it affect other electric charges. The quantitative experiments of the investigators mentioned showed that the charge on condensers of different geometrical form to which a given difference of potential was applied always changed in a certain ratio when any assigned insulating material was replaced by any other assigned insulating material. Different insulators were therefore said to be characterized by different "specific inductive capacities." The specific inductive capacity of a vacuum is now universally adopted as a standard, and its value is put equal to unity on the electrostatic system of units. The specific inductive capacity of air only differs very slightly from it. In this book, for reasons which will appear, we shall use the term dielectric coefficient instead of specific inductive capacity.

The potential difference between two electric charges is, by its definition, determined by the distribution of the electric intensity in the field surrounding them. It is clear, therefore, that these experiments prove that the forces between charged bodies are not determined solely by the magnitude of their charges and their distance apart. Those of our previous results which are based upon the law of the inverse square will therefore require modification if we are to account for the behaviour of charged bodies in the neighbourhood of insulating materials.

The relation $V=\int E ds$ is independent of the law of force, as is also the device of mapping out the field by means of tubes generated by the lines of electric force. The ideas underlying them can therefore be applied even when dielectric media are present. We shall suppose that a tube of force starts from each element of area which embraces a unit of positive charge. These tubes must either end on another charge or flow on to infinity. This follows since two different equipotential surfaces cannot intersect and since the lines of force are at right angles to the equipotential surfaces. It is also necessary that the charge at the negative end of a tube should be equal and opposite to the charge at the positive end. This is required by the fact that the two sides of a condenser acquire equal and opposite charges whatever the nature of the intervening medium.

The tubes which are determined in this way we shall call tubes of induction and we shall define the induction D at any point as the number of tubes which cross a unit area drawn perpendicular to the direction of the tubes at that point.

The induction is a vector quantity and its components are given by the usual rule for the resolution of vectors. Thus if D_θ is the component of D along a line OQ making an angle θ with the direction OP of the resultant D, it is clear that the number of tubes which cross unit area perpendicular to OQ is

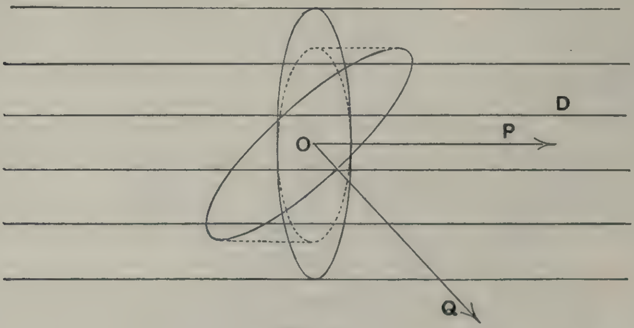


Fig. 7.

the same as that of those which cross an area $\cos\theta$ perpendicular to OP, so that

$$D_\theta = D\cos\theta.$$

It is evident that the cross-sectional area of the tube of induction passing through any point is the inverse of the induction at that point.

Gauss's Theorem.

We are now in a position to consider the form which Gauss's Theorem takes when the field embraces dielectric media. Consider the value of the integral $\iint D_n dS$ taken over any closed surface, where D_n is the component of the induction along the outward drawn normal at an element dS of the surface. If OPQ represents the direction of the tubes crossing dS, and if the angle $RPQ = \theta$, then $D_n = D\cos\theta$, where D is the resultant induction at dS. Now $dS\cos\theta$ is the projection of dS on a plane perpendicular to the direction of D, so that

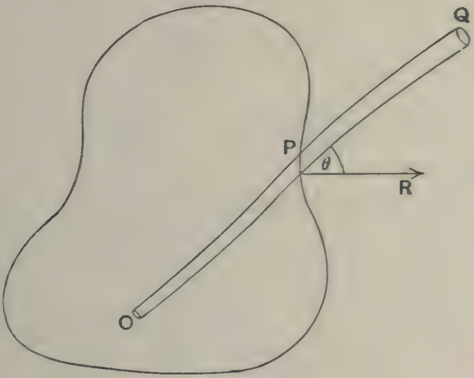


Fig. 8.

$$D \times dS\cos\theta = D_n dS$$

is the number of tubes which cross dS from the inside to the outside of the surface. When the tubes cross from the outside to the inside of the closed surface $D_n dS$ will have a negative value. Let us first consider the effect of those elements for which $D_n dS$ has a positive value. Each of the tubes crossing them will start from a unit positive charge beyond dS on the internal side. This unit charge may either be without or within the surface. If it is without, the tube will first cross the surface at some other point, and it will give rise to an equal and opposite element $D_n'dS'$. The total contribution of such tubes to the integral will be zero. Each tube crossing an element dS in the positive direction and having its origin within the surface will start on a unit positive charge, so that the total contribution of all the elements $D_n dS$ which are positive will be equal to that part of the positive charge within the surface which gives rise to tubes of force which do not end on a negative charge inside. In a precisely similar manner we can show that the total contribution of all the elements $D_n dS$ which are negative is equal to that part of the negative charge within the surface on which tubes of force end which do not start from positive charges inside the surface.

It follows that the value of $\iint D_n dS = \iint D \cos\theta \, dS$ taken over any closed surface is equal to the algebraic sum e of all the charges enclosed by the surface.

We thus see that Gauss's Theorem can be extended to dielectric media provided the normal electric intensity is replaced by the normal induction.

By applying Gauss's Theorem to a cylindrical region bounded by a tube of force and by two equipotential surfaces, one just inside and the other just outside a conductor, we find that the induction just outside the surface of a conductor is along the normal to the surface and equal to the charge per unit area of it. This is the general form of Coulomb's Law.

Poisson's and Laplace's Equations.

Let us now apply Gauss's Theorem to the rectangular parallelepiped whose angular points are the combinations of $x \pm \frac{1}{2}dx$, $y \pm \frac{1}{2}dy$, $z \pm \frac{1}{2}dz$. The induction at the centre of the element being $D\,(D_x,\ D_y,\ D_z)$, the mean outward normal induction over the faces perpendicular to the axis of x will be $-\left(D_x - \frac{\partial D_x}{\partial x}\frac{dx}{2}\right)$ and $+\left(D_x + \frac{\partial D_x}{\partial x}\frac{dx}{2}\right)$ by Taylor's Theorem. The corresponding fluxes of induction are

$$-\left(D_x - \frac{\partial D_x}{\partial x}\frac{dx}{2}\right) dy\,dz$$

and

$$+\left(D_x + \frac{\partial D_x}{\partial x}\frac{dx}{2}\right) dy\,dz.$$

The total flux of induction over this pair of faces is therefore $\frac{\partial D_x}{\partial x} dxdydz$. Treating the other two pairs of faces similarly, we see that the total flux of induction over all the six faces is

$$\left\{\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}\right\} dxdydz.$$

This is equal to the charge inside, which is $\rho dxdydz$. We therefore get

$$\operatorname{div} D=\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}=\rho \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

as the general form of Poisson's Equation. The general form of Laplace's Equation follows if we put $\rho=0$.

Induction and Electric Intensity.

From the definition of the induction it follows that in geometrically identical systems of conductors, furnished with identical distributions of electric charge, the induction will be identical at every point whatever the nature of the intervening dielectric, provided it is homogeneous and isotropic. The experiments of Cavendish and Faraday show that the intensities at corresponding points are inversely as the dielectric coefficient, since for geometrically identical systems the forces are as the differences of potential. It follows that the relation between the induction and the electric intensity at any point in a dielectric medium is

$$D=\kappa E \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2),$$

where κ is the dielectric coefficient. Since $E=-\operatorname{grad} V$, we may also write the preceding results in the following forms:

Gauss's Theorem.

$$\iint \kappa \frac{\partial V}{\partial n} dS=-\iiint \rho\, d\tau \quad \ldots\ldots\ldots\ldots\ldots\ldots(3).$$

Coulomb's Law.

$$\kappa \frac{\partial V}{\partial n}=-\sigma \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4).$$

Poisson's Equation.

$$\frac{\partial}{\partial x}\left(\kappa \frac{\partial V}{\partial x}\right)+\frac{\partial}{\partial y}\left(\kappa \frac{\partial V}{\partial y}\right)+\frac{\partial}{\partial z}\left(\kappa \frac{\partial V}{\partial z}\right)=-\rho \quad \ldots\ldots(5),$$

or, when κ is constant throughout the space under consideration,

$$\kappa \nabla^2 V=-\rho \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6).$$

The Energy in the Field.

The theorem (equation (17), Chap. II) that the potential energy W of any system of charges is equal to

$$\tfrac{1}{2}\iiint V\rho\, d\tau + \tfrac{1}{2}\iint V\sigma\, dS$$

does not depend on the law of force between two elements of charge, and is therefore true when dielectric media are present. Since $\rho = \operatorname{div} D$ and $\sigma = D_n$, we have

$$W = \tfrac{1}{2}\iiint V \operatorname{div} D\, d\tau + \tfrac{1}{2}\iint V D_n\, dS.$$

But
$$\iiint V\frac{\partial D_x}{\partial x}\, d\tau = \iiint \frac{\partial}{\partial x}(VD_x)\, d\tau - \iiint D_x \frac{\partial V}{\partial x}\, d\tau$$
$$= -\iint V l D_x\, dS - \iiint D_x\frac{\partial V}{\partial x}\, d\tau,$$

where l is the x direction cosine of the normal to dS. Since

$$D_n = lD_x + mD_y + nD_z,$$

we see, by similarly integrating the other components of $\iiint V \operatorname{div} D\, d\tau$ by parts, that

$$\iiint V\operatorname{div} D\, d\tau = -\iiint \left\{D_x\frac{\partial V}{\partial x} + D_y\frac{\partial V}{\partial y} + D_z\frac{\partial V}{\partial z}\right\} d\tau - \iint V D_n\, dS.$$

Hence
$$W = -\frac{1}{2}\iiint \left\{D_x\frac{\partial V}{\partial x} + D_y\frac{\partial V}{\partial y} + D_z\frac{\partial V}{\partial z}\right\} d\tau \quad \text{......(7).}$$

The energy per unit volume of the field is therefore

$$-\frac{1}{2}\left\{D_x\frac{\partial V}{\partial x} + D_y\frac{\partial V}{\partial y} + D_z\frac{\partial V}{\partial z}\right\}.$$

If E_x, E_y, E_z are the components of E, since $E = -\operatorname{grad} V$, this may be written

$$\tfrac{1}{2}\{D_xE_x + D_yE_y + D_zE_z\} \quad \text{..................(8).}$$

The sum of the products of the components of two vectors taken in this way is called the scalar product of the vectors*. It is often written (DE), so that in this notation the energy of unit volume of the medium

$$= \tfrac{1}{2}(DE) \quad \text{..........................(9).}$$

---

* See Webster, *Electricity and Magnetism*, Chap. I.

In isotropic media D and E are in the same direction and $D=\kappa E$, so that the energy per unit volume of the medium

$$=\frac{1}{2}\kappa E^2=\frac{1}{2}\frac{D^2}{\kappa} \quad \text{.....................(10).}$$

This reduces to the result in Chapter II when $\kappa=1$.

Conditions at the Boundary between Two Dielectric Media.

It is very important to determine how the induction and the electric intensity change as we pass across the surface separating two different insulating media. Consider first of all an element of area dS of the surface, so small that its curvature may be neglected. Describe the prism generated by lines passing through the boundary of dS and at right angles to its plane. Let the prism be terminated by planes parallel to dS, and let the height of the prism be a small quantity of the second order, if the width of dS is of the first order. Then one of the ends of the flat prism thus constructed will be in the first medium and the other in the second, whilst the sides are partly in one medium and partly in the other. Now apply Gauss's theorem to the prism. The induction is necessarily finite, so that the normal flux over the sides of the prism vanishes, since their area is negligible compared with that of the ends. Let D_1 be the magnitude of the resultant induction at the boundary in the first medium and let it make an angle θ_1 with the normal, D_2 and θ_2 being the corresponding quantities in the second medium. The areas of the two ends being each equal to dS in the limit, the flux of normal induction over the first will be $D_{n_1}dS=D_1\cos\theta_1 dS$ and over the second $D_{n_2}dS=-D_2\cos\theta_2 dS$. The sum of these two is equal to the total charge inside the prism, which is σdS, if σ is the charge per unit area of the boundary. We see therefore that the induction at the boundary between two media will always satisfy the relation

$$D_1\cos\theta_1-D_2\cos\theta_2=\sigma.$$

In the majority of cases there will be no charge on the boundary surface, and thus the component of the induction normal to the surface will have the same value on both sides of the boundary. In other words:—when there is no charge on the surface of separation the normal component of the induction is continuous from one medium to the other; when this is not the case it is discontinuous by an amount equal to the charge per unit area of the surface.

A similar relation is satisfied by the *tangential* component of the electric *intensity*. Consider the rectangle which forms a central section through the axis of the prism just considered. Let us calculate the work done in taking a unit electric charge round the sides of such a rectangle. The force being finite everywhere, the work done along the short sides vanishes in comparison with that done along the long sides. The long sides are of equal length s in the limit, and if T_1, T_2 are the values of the tangential components of the electric intensity in the two media, the work done in taking a unit charge round the rectangle will be $T_1 s - T_2 s$. This must vanish, otherwise we could obtain an indefinite amount of work by repeating the operation an indefinite number of times. We therefore conclude that $T_1 = T_2$, or, in other words, that the component of electric intensity tangential to the surface is continuous in passing from one medium to the other. It is clear that this result holds whether there is a charge on the boundary or not.

These results enable us to obtain the law of refraction of the tubes of induction at the boundary between the two media. Let us suppose that the interface is uncharged and let κ_1, κ_2 be the dielectric coefficients of the two media. Then, since the normal component of the induction is continuous, we have

$$D_1 \cos\theta_1 = D_2 \cos\theta_2,$$

and, since the tangential electric intensity is continuous,

$$\frac{D_1}{\kappa_1} \sin\theta_1 = \frac{D_2}{\kappa_2} \sin\theta_2,$$

whence

$$\tan\theta_1 / \tan\theta_2 = \kappa_1/\kappa_2 \quad \text{.................}(11),$$

or the tangents of the angles which the tubes of induction make with the normal to the surface are directly as the dielectric coefficients of the media.

We see from the results of the last two sections that in dealing with problems involving dielectric media the induction must be a solution of the equation $\operatorname{div} D = \rho$. The further condition has to be satisfied that at the surface separating any two media the components of the induction normal to the surface differ on the two sides of it, by an amount equal to the charge per unit area of the surface at that point. In addition, the tangential component of the electric intensity must have the same value on both sides of the surface. There is only one solution of the differential equation which satisfies the conditions in any assigned case; so that any particular solution of the differential equation which can be made to satisfy the boundary conditions will determine the field uniquely. The boundary conditions which we have established for dielectric media include those relating to conductors in a vacuum, or in any dielectric, as particular cases.

Poisson's Theory of Dielectric Media.

There is a certain view of the behaviour of dielectric media the mathematical development of which is largely due to the French physicist Poisson. According to this view the modification of the electric field arising from the presence of dielectrics is due to the substance of the dielectric being thrown into a peculiar electrical condition by the external field. This condition arises from the displacement of electricity in the ultimate particles of the medium in such a way that each particle acquires a positive charge at one end and a negative charge at the other. When this occurs the medium is said to be polarized and we shall see that the polarization is measured by the product of the displacement and the charge per unit volume.

We shall first of all consider the nature of polarization as it presents itself from the point of view which regards electricity as continuously distributed, after the manner of a fluid, throughout all bodies. Consider two equal spheres, and let one of them be filled with a uniform distribution of positive electricity and the other with a distribution of negative electricity, identical with the first except for the difference of sign. Imagine the two spheres coincident in position; we shall then have an uncharged body which will give rise to no electrical effects. Now suppose that, whilst one of the spheres is fixed, every point of the other is given a certain equal displacement, so that this sphere moves a small distance after the manner of a rigid body. The region of space where the spheres overlap will still be free from electric charge, but there will be a layer of positive electricity over the surface of the sphere on that side towards which the positive electricity has been displaced, and a similar distribution of negative electricity over the opposite face. The thickness of the layer measured parallel to the axis of displacement will be the same at every point of the surface, so that the thickness measured parallel to the normal at any point is equal to the magnitude of the displacement multiplied by the cosine of the angle that the normal makes with the direction of the displacement. This result will clearly hold whatever the shape of the body may be. By making the density of the charges indefinitely big and the displacement indefinitely small, in such a way that their product remains finite, the same result may be obtained without altering the shape of the body. A body whose electrical behaviour can be represented in this way is said to be uniformly polarized. The direction of the axis of polarization is the same as that of the relative displacement of the positive to the negative distributions. To facilitate discussion the charges which are imagined to arise in the dielectric in this way will be referred to as "fictitious" in contradistinction to the "real" charges which occur for example on the surface of conductors. The legitimacy of this distinction will be considered more fully later.

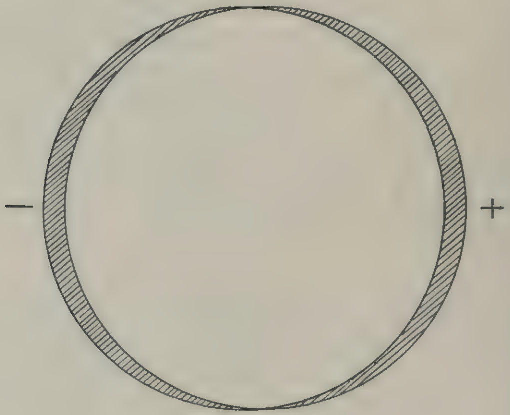


Fig. 9.

Another way of regarding polarization is to suppose the body divided into equal cellular elements with the axis of each element parallel to the direction of the polarization. To represent uniform polarization we then suppose each cell to develop a positive charge on one face and an equal and opposite charge on the opposite face, the faces affected being those normal to the axis of polarization.

A body constructed in this way is identical electrically with a body made up of two uniform distributions of positive and negative electricity which have undergone relative displacement. This is clear, because in the interior of the body equal and opposite charges coincide at the contiguous surfaces of opposite elements, so that there is no volume charge at any point; whilst at the external boundary there is a charge, positive at one end of the body and negative at the other end, whose magnitude is the same for each portion of the surface in which it is intersected by the sides of a cell. But the area of this intersection is inversely as the cosine of the angle that the normal to the surface makes with the axis of the cells. Thus the density of this fictitious surface layer varies as the cosine of the angle that the normal makes with the axis of polarization, and the distribution is completely identical with that obtained by displacing originally coincident electric charges.

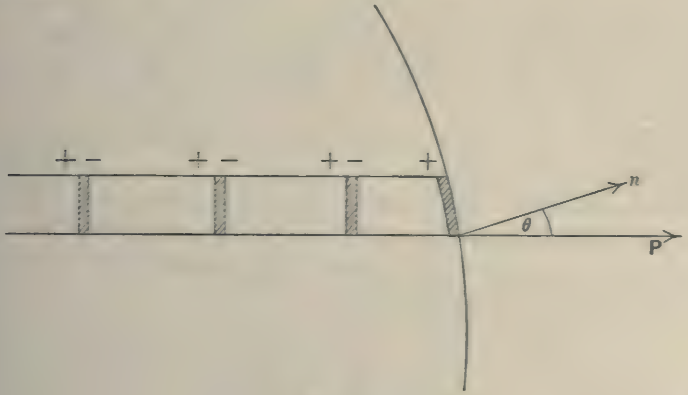


Fig. 10.

In some problems it is more convenient to regard polarized media as displaced charges which were originally coincident, whilst in others the point of view which considers them as made up of cellular elements having opposite charges on opposite faces has advantages.

If we suppose a polarized body to be intersected by any surface, the two resulting portions of the body will still be polarized. It is necessary therefore that there should be developed at any such surface of separation a double distribution of electric charge, of equal amount and opposite sign at any point of the interface.

This development is required in order to make the algebraic sum of all the charges on each part of the body zero. We see that the cellular elements which we have considered previously can be regarded as arising in this way by a series of fractures of the body along surfaces parallel and perpendicular to the axis of polarization at any point.

So far we have not defined polarization in a manner sufficiently precise to admit of numerical expression. We shall now define the intensity of polarization, or the polarization, as we shall, for the sake of brevity, usually call it, as the electric moment per unit volume of the medium at any point. Consider one of our cellular elements of length δx; let the area of the normal end faces be δS and let $\pm\sigma$ be the surface density of the charge over them. Then the charge on each end is $\pm\sigma\delta S$ and the electric moment of the element is $\sigma\delta x\delta S=\sigma\delta\tau$, where $\delta\tau$ is the volume of the element. The electric moment is therefore proportional to the volume of the element, and the electric moment per unit volume is equal to the surface density σ. Clearly σ is also equal to the charge which develops per unit area over an intersection of the body at the point under consideration by a plane perpendicular to the axis of polarization. Thus the polarization may be defined either as the electric moment per unit volume of the body or as the charge per unit area of an interface perpendicular to the direction of polarization. This interface may include the external surface of the polarized body as a special case.

The statement in the preceding paragraph is to be regarded as the definition of the resultant polarization. The polarization, however, is a vector quantity and this aspect of the case can be provided for by a slight modification of the definition. We now define the polarization in any direction at any point as the moment per unit volume of a thin slab of the polarized medium described about the given point and with its faces perpendicular to the given direction. The slab can conveniently be regarded as made up of a series of the prismatic elements already considered (see Fig. 11). If ΔS is the area of each face of the slab the charge on them will clearly be $\pm\sigma\Delta S\cos\theta$, where θ is the angle between the normal to the slab and the direction of the resultant polarization. The perpendicular distance between the two layers of electric charge is $\cos\theta\,\delta x$, if δx is the length of a cell. The electric moment (perhaps strength would be a better term) of the slab is therefore $\sigma\cos^2\theta\,\Delta S\,\delta x$. The volume of the slab is

$$\cos\theta\,\Delta S\,\delta x,$$

so that the polarization along the direction normal to the slab is clearly

$$P_\theta = \sigma\cos\theta = P\cos\theta \quad \ldots\ldots\ldots\ldots\ldots\ldots(12),$$

where P is the resultant polarization. The components of the polarization are therefore obtained by the usual rule for the resolution of vectors.

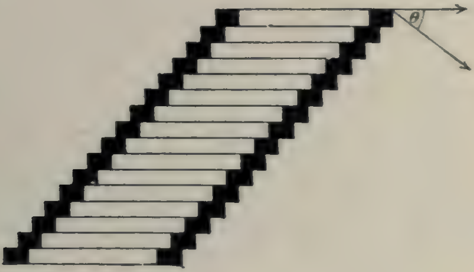


Fig. 11.

Polarization, Induction and Electric Intensity.

We are now in a position to determine the relation between these three vectors. Let us first consider the nature of the electric field between the plates of a parallel plane condenser filled with a dielectric of specific inductive capacity κ. Then the induction D and the electric intensity E are both normal to the plates and $\kappa E = D = \sigma$, where σ is the charge per unit area of the plates.

We shall make the hypothesis that the polarization is caused by the electric intensity in the dielectric, and that the two vectors are coincident in direction. This standpoint will be found to be fully justified when we come to consider the phenomena from the point of view of the electron theory, according to which the polarization arises from an actual displacement of the ultimate electrified particles in the direction of the field. The hypothesis is undoubtedly true for isotropic dielectric substances. In the case of crystalline substances the electric intensity does not in general coincide in direction with the polarization, and a similar statement holds with regard to the magnetic behaviour of crystalline media. These exceptions are due to complications which it would lead us too far out of our way to discuss, and their existence does not, in any way, vitiate the general principles involved.

Now on the theory of polarized media the electric intensity is supposed to arise partly from the external charged bodies and partly from the charges arising from the polarization at the boundary of the medium. In the present case the amount of this charge will be ± P per unit area of the plates. It will be negative where σ is positive and vice versa. We therefore have

$$E = \sigma - P = D - P \quad\text{..................(13)},$$

and

$$\kappa E = D;$$

whence

$$P = (\kappa - 1) E = \frac{\kappa - 1}{\kappa} D \quad\text{................(14)}.$$

As we shall see (p. 57) these relations between P, E and D are perfectly general, although we have only deduced them from a very special case, that of a parallel plate condenser. E and D are equal to the forces which would be exerted on a unit charge at a point in cavities of certain shapes made in the dielectric. Taking the case of E first, we observe that E is the force which would be exerted on a unit positive charge, placed at the point where E is measured, by the so-called real charges in the field together with the Poisson distribution to which the polarization of the dielectric medium is equivalent. This will be the actual force on a unit charge in an actual cavity of indefinite length and infinitesimal cross section, whose axis follows the direction of the lines of force at every point. For the charges which develop on the walls of such a cavity, owing to the existence of polarization, will be confined to the two ends, and the contribution from these to the force inside the cavity will vanish in the limit, when the cross section is made to diminish indefinitely. The force will therefore be determined solely by the real and fictitious charges in the field and will be identical with E.

Next consider a cavity whose cross section is great compared with its length, although both are infinitesimal. Let the end faces of this cavity be normal to the direction of the polarization. Then there will be, on each of the faces, a distribution of electric charge due to the polarization and equal to $\pm P$ per unit area. The force on a unit charge in such a cavity will arise partly from the charge on the walls and partly from the charges, real and fictitious, in the rest of the field. The former part is equal to P and the latter to E. The total force is therefore $P + E = D$. The resultant induction D at any point is therefore equal to the force which would act on a unit positive charge placed in a very flat cavity cut perpendicular to the direction of the lines of force at that point.

It is easy to show that the component of the induction in any direction is the component, in that direction, of the force which would be exerted on a unit positive charge placed in a similar flat cavity with its end faces normal to the direction in question.

The foregoing specifications of the induction and electric intensity satisfy the conditions which we have already laid down for them (p. 45) at the interface between two media K_1 and K_2 of different dielectric coefficients. For, consider two flat cavities A and B parallel to the interface C and indefinitely near to it. The normal component of the induction just inside K_1 will be equal to the force on a unit charge in A and that just inside K_2 will be equal to the force acting on a similar charge in B. If the resultant intensities in the two media are E_1 and E_2, and the resultant polarizations are P_1 and P_2, and if they make angles θ_1 and θ_2 with the normals to the surface, then it follows from the results of the preceding paragraph that the normal force in A is $(E_1 + P_1)\cos\theta_1$, and that in B is $(E_2 + P_2)\cos\theta_2$. That these forces are equal to one

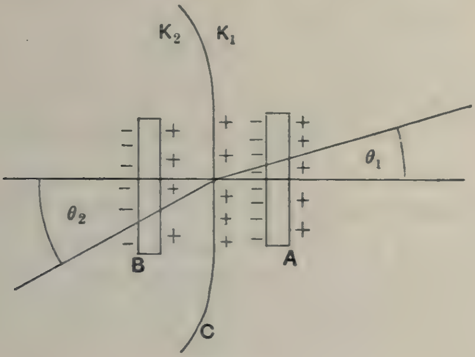


Fig. 12.

another can be seen by considering the charges, real and fictitious, from which they arise. The force inside A arises from

(1) charges at a distance,

(2) the polarization charges on the ends of A,

(3) the fictitious charge over the boundary C.

The last is made up of a positive charge arising from the polarization of K_1 and a different negative charge arising from the polarization of K_2. The positive charge on C is equal to the negative charge on an equal area of the left face of A, and the negative charge on C is equal to the positive charge on an equal area of B. The total force in A is thus the same as that which would arise from

(1) charges at a distance,

(2) the positive charge on the right of A,

(3) the negative charge on the left of C.

In a similar way we can show that the force in B arises from

(4) charges at a distance,

(5) the positive charge on the right of C,

(6) the negative charge on the left of B.

But (1) = (4), (2) = (5) and (3) = (6). It follows that our specification of the induction makes its normal component continuous.

The tangential component of the electric intensity is continuous in crossing the boundary, since the only change in the electric intensity which occurs is that which arises from the change in position relative to the fictitious charge on the boundary, and this gives rise only to a force normal to the interface.

Variable Polarization.

So far we have confined our attention to uniformly polarized media, i.e. to cases in which the polarization P has the same magnitude and direction at all points. The number of such cases is strictly limited, and it is very important to study the behaviour of media in which the polarization varies from point to point. Bodies polarized in this way are generated by the sliding of oppositely charged coincident distributions, if the original distributions of electric charge are not uniform. We shall, however, look at the matter from the point of view that the polarization arises from the development of charges on the faces of the ultimate elements of the body in the manner which we have already considered. Consider any elementary rectangular parallelepiped of the material whose sides are δx, δy, δz, and the coordinates of whose centre are x, y, z. Let the resultant polarization at the centre be P and let its components parallel to the coordinate axes be P_x, P_y and P_z. The charges which develop over the faces of the parallelepiped owing to the polarization $\boldsymbol{P}$ of the parallelepiped itself will be

$$\pm P_x \delta y \delta z, \ \pm P_y \delta x \delta z, \text{ and } \pm P_z \delta x \delta y$$

respectively. These are the mean values taken over the whole of each respective face of the parallelepiped. If we consider any one of these faces, for instance that which is determined by $x = x + \frac{1}{2}\delta x$, we see that the next element, in this case the one to the right, will give rise to a charge over it $= -\left(P_x + \frac{\partial P_x}{\partial x}\delta x\right)\delta y \delta z$, since $P_x + \frac{\partial P_x}{\partial x}\delta x$ is the mean value of the x component of the polarization in the next element of volume. This face is therefore to be regarded as carrying a charge due to polarization equal altogether to $-\frac{\partial P_x}{\partial x}\delta x \delta y \delta z$. One half of this is to be considered as belonging to the next element of volume, so that only $-\frac{1}{2}\frac{\partial P_x}{\partial x}\delta x \delta y \delta z$ belongs to the element under consideration. The face for which $x = x - \frac{1}{2}\delta x$ gives rise to an equal amount, so that the total charge arising from the faces perpendicular to the axis of x is $-\frac{\partial P_x}{\partial x}\delta x \delta y \delta z$. In a similar manner we can show that the pairs of faces perpendicular to Oy, Oz carry charges which contribute

$$-\frac{\partial P_y}{\partial y}\delta x \delta y \delta z \text{ and } -\frac{\partial P_z}{\partial z}\delta x \delta y \delta z$$

respectively. Thus when the medium is polarized non-uniformly, there is associated with each element of volume a charge equal to

$$-\left(\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z}\right)\delta x \delta y \delta z \ldots\ldots\ldots\ldots\ldots(15).$$

This charge is a fictitious charge like the charge which develops on the surface of dielectric substances on the polarization theory of the behaviour of dielectric media. It is to be definitely distinguished from the true charge carried by conductors and other electrically charged bodies in the field.

We shall now consider the relation between the fictitious volume and surface charges which arise in a polarized body when the polarization is not uniform. Denoting the fictitious charges by dashes we have

$$\iiint \rho' \,dx\,dy\,dz = -\iiint \left(\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z}\right) dx\,dy\,dz.$$

We can evaluate the integral on the right, taken throughout the polarized body, by the method of Green's Theorem. If l, m, n are the direction cosines of the external normal at any point, we see that

$$\iiint \frac{\partial P_x}{\partial x}\,dx\,dy\,dz = \iint l P_x \,dS,$$

with similar expressions for the remaining constituents of the integral. We therefore have

$$\begin{aligned}\iiint \rho' \,dx\,dy\,dz &= -\iint (lP_x + mP_y + nP_z)\,dS \\ &= -\iint P_n \,dS \quad\text{.........................(16)},\end{aligned}$$

where P_n is the component of the polarization along the outward normal to the surface at dS. This is equal to σ', the surface density of the fictitious charge arising from the polarization at that point. Hence

$$\iiint \rho' \,dx\,dy\,dz + \iint \sigma' \,dS = 0 \quad\text{..............(17)},$$

or the algebraic sum of the volume and surface charges arising from polarization is zero. This result would have been obvious had we developed the properties of non-uniformly polarized media by the sliding of originally coincident equal and opposite distributions of charge.

According to the polarization theory of dielectric media the field of force which arises when such media are present is the same as that which would arise if the medium were all aether, provided the true charges ρ and σ are accompanied by the fictitious charges ρ' and σ' of the polarized distribution. The electric intensity will therefore obey the form of Poisson's equation which obtains in the free aether, provided the density at any point is supposed to be equal to $\rho + \rho'$. We therefore have

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho + \rho' \quad \text{..................(18).}$$

Now

$$\rho' = -\left(\frac{\partial P_x}{\partial x} + \frac{\partial P_y}{\partial y} + \frac{\partial P_z}{\partial z}\right),$$

so that

$$\frac{\partial}{\partial x}(E_x + P_x) + \frac{\partial}{\partial y}(E_y + P_y) + \frac{\partial}{\partial z}(E_z + P_z) = \rho \quad \text{....(19).}$$

Comparing this with equation (1), we see that the vector whose components are $E_x + P_x$, $E_y + P_y$, $E_z + P_z$ is identical with the induction D, so that the identification from a particular case on p. 52 is perfectly general.

By integrating both sides of equation (19) throughout any enclosed volume we see that the true charge inside is equal to $\iint D_n dS$ over the bounding surface, in agreement with p. 42.

The Fictitious Charges on the Electron Theory.

Although it is necessary, in discussing the results of electrostatic experiments, to distinguish between "true" charges like those which are communicated from a conductor to the plates of a condenser and the "fictitious" charges which appear to reside in the dielectric, there is no very profound difference between them. According to the electron theory one is just as true a charge as the other, although its reality is not so readily made obvious by experiment. The electron theory regards a dielectric as a certain type of distribution of electrons in space, and in this space the true electric intensity satisfies the equation div $E = \rho$. This equation is assumed to be true when the element of volume is a small part of an electron. When the element of volume is enlarged so as to contain a great many electrons the equation will become div $\bar{E} = \bar{\rho}$, where the bars denote average values. Thus $\bar{\rho}$ is equal to the $\rho + \rho'$ of equation (18) and ρ' is just as real a part of the average density of electrification as ρ. This point will be considered more fully in the sequel.

CHAPTER IV

THE ELECTRON THEORY OF DIELECTRIC MEDIA

Potential due to a Doublet.

CONSIDER the doublet formed by equal positive and negative charges $\pm e$ placed at Q and O respectively, where $OQ = s$. Let V be the potential at a distant point P, where $QP = r$ is large compared with s. Then

$$4\pi V = \frac{e}{QP} - \frac{e}{OP} = e\frac{OQ\cos\theta}{QP^2}$$

$$= \frac{es}{r^2}\cos\theta = \frac{\mu}{r^2}\cos\theta \quad \text{.....................(1)},$$

in the limit when s vanishes compared with r. μ is called the moment of the doublet OQ.

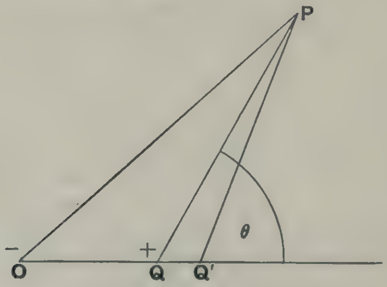


Fig. 13.

This result may be written rather differently. Let the axis of the doublet be given a small displacement parallel to its length, so that Q moves to Q', where $QQ' = \delta s$. We have

$$PQ'^2 = PQ^2 + QQ'^2 - 2PQ\,.\,QQ'\cos\theta,$$

or $$(r + \delta r)^2 = r^2 + \delta s^2 - 2r\,.\,\delta s\,.\cos\theta,$$

so that to the first order

$$2r\delta r = -2r\delta s\cos\theta,$$

and

$$\cos\theta = -\frac{\partial r}{\partial s};$$

$$\therefore\ 4\pi V = -\frac{\mu}{r^2}\frac{\partial r}{\partial s} = \mu\frac{\partial}{\partial s}\left(\frac{1}{r}\right) \quad \text{..............(2)},$$

where $\frac{\partial}{\partial s}$ denotes differentiation along the direction of the positive axis of the doublet. In this differentiation the moment of the doublet is supposed to be constant, so that the increment δs will be determined by the displacement of the centre of the doublet. Let the coordinates of the centre be a, b, c, then s is to be regarded as a function of a, b and c; whence it follows that

$$V = \frac{\mu}{4\pi}\left\{\frac{\partial}{\partial a}\left(\frac{1}{r}\right)\frac{\partial a}{\partial s} + \frac{\partial}{\partial b}\left(\frac{1}{r}\right)\frac{\partial b}{\partial s} + \frac{\partial}{\partial c}\left(\frac{1}{r}\right)\frac{\partial c}{\partial s}\right\} \quad \text{......(3)}.$$

Now $\frac{\partial a}{\partial s}, \frac{\partial b}{\partial s}, \frac{\partial c}{\partial s}$ are the direction cosines l, m, n of the axis s of the doublet. So that

$$V = \frac{\mu}{4\pi}\left\{l\frac{\partial}{\partial a}\left(\frac{1}{r}\right) + m\frac{\partial}{\partial b}\left(\frac{1}{r}\right) + n\frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} \quad \text{.........(4)}.$$

If we resolve the moment μ of the doublet into components parallel to the axes, these will be $\mu_1 = l\mu$, $\mu_2 = m\mu$, $\mu_3 = n\mu$ and

$$V = \frac{1}{4\pi}\left\{\mu_1\frac{\partial}{\partial a}\left(\frac{1}{r}\right) + \mu_2\frac{\partial}{\partial b}\left(\frac{1}{r}\right) + \mu_3\frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} \quad \text{......(5)}.$$

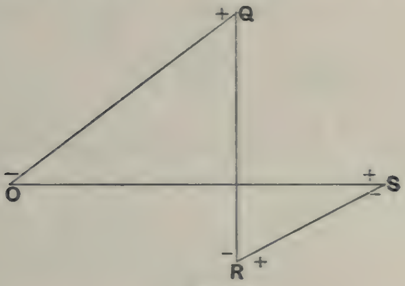


Fig. 14.

Thus the potential, and therefore also the field of force, arising from the doublet is the same as that due to the sum of the effects of its components. This result is at once obvious geometrically (see Fig. 14). For the resolution of the doublet OQ into its components OS, SR and RQ parallel to the axes is equivalent to the placing of two equal and opposite charges at each of the points *R* and *S*. This is clearly incapable of changing the field in any way, so that the field due to the three components must be the same as that due to the original doublet.

Potential due to Polarization.

We have seen that a polarized medium can be regarded as being built up of a series of cells having equal and opposite charges spread over opposite faces. When the axis of the cells is parallel to the direction of the polarization, only one pair of the faces, those which are normal to the direction of the polarization, will be charged, and each cell will behave like a single doublet. The moment of this doublet, being equal to the product of the length of the cell by the charge on its end faces, is clearly equal to the product of the resultant polarization by the volume of the cell. When the faces of the cells, supposed to intersect orthogonally, do not bear any simple directional relation to the axis of polarization, charges will develop over all the faces of the cells. These will be equal and opposite for each opposite pair of faces, and will be equal in magnitude to the area of each face multiplied by the component of polarization normal to it. The moment of the doublet to which the cell is equivalent will thus be equal to that of the doublet whose components are the components of the polarization normal to the faces of the cell, each multiplied by the volume of the cell.

Since every element of volume of a polarized medium is equivalent to a doublet, this result enables us to write down the expression for the potential arising from an element of volume of a polarized medium. Let a, b, c be the coordinates of the centre of the element of volume, its sides being equal to da, db, dc. Let l, m, n be the direction cosines of P, the resultant polarization. Then the polarized element is equivalent to a doublet, the components of whose moment are $(lP, mP, nP)\,da\,db\,dc$. The contribution of this element to the potential at a distant point x, y, z is therefore

$$\frac{P}{4\pi}\left\{l\frac{\partial}{\partial a}\left(\frac{1}{r}\right)+m\frac{\partial}{\partial b}\left(\frac{1}{r}\right)+n\frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\}da\,db\,dc$$

$$=\frac{1}{4\pi}\left\{P_x\frac{\partial}{\partial a}\left(\frac{1}{r}\right)+P_y\frac{\partial}{\partial b}\left(\frac{1}{r}\right)+P_z\frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\}da\,db\,dc \quad \ldots(6),$$

where $$r^2=(x-a)^2+(y-b)^2+(z-c)^2.$$

It follows that the potential arising from the whole of the polarized medium is

$$V = \frac{1}{4\pi}\iiint\left\{P_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) + P_y \frac{\partial}{\partial b}\left(\frac{1}{r}\right) + P_z \frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} da\,db\,dc \quad \ldots(7).$$

In these expressions P_x, P_y and P_z are the components of the polarization at the point a, b, c, and not at x, y, z.

The preceding formula can also be obtained by a transformation of the usual expression for the potential of a distribution of electric charge

$$V = \frac{1}{4\pi}\iiint \frac{\rho}{r}\,d\tau + \frac{1}{4\pi}\iint \frac{\sigma}{r}\,dS.$$

If this is applied to the case of a polarized medium, ρ and σ will represent what we have called the fictitious charges of polarization. Thus

$$\rho = -\left(\frac{\partial P_x}{\partial a} + \frac{\partial P_y}{\partial b} + \frac{\partial P_z}{\partial c}\right),$$

and $$\sigma = -\left(P_x \cos n_i a + P_y \cos n_i b + P_z \cos n_i c\right),$$

where $\cos n_i a$, $\cos n_i b$, $\cos n_i c$ denote the direction cosines of the internal normal to the bounding surface, referred to axes parallel to a, b, c. Thus

$$V = -\iiint \frac{1}{4\pi r}\left(\frac{\partial P_x}{\partial a} + \frac{\partial P_y}{\partial b} + \frac{\partial P_z}{\partial c}\right) da\,db\,dc$$
$$- \iint \frac{1}{4\pi r}\left(P_x \cos n_i a + P_y \cos n_i b + P_z \cos n_i c\right) dS.$$

Integrating the volume integral by parts, we have

$$\iiint \frac{1}{r}\frac{\partial P_x}{\partial a}\,da\,db\,dc = \iiint \frac{\partial}{\partial a}\left(\frac{1}{r}P_x\right) da\,db\,dc - \iiint P_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) da\,db\,dc$$
$$= -\iiint P_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) da\,db\,dc - \iint \frac{P_x}{r}\cos n_i a\,dS.$$

Since similar expressions are obtained from the other two terms of the integral it follows that

$$V = \frac{1}{4\pi}\iiint\left\{P_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) + P_y \frac{\partial}{\partial b}\left(\frac{1}{r}\right) + P_z \frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} da\,db\,dc.$$

Polarized Shells.

A polarized shell is a superficial distribution of polarization. It may be regarded as a region bounded by two surfaces at an infinitesimal distance apart and carrying opposite charges on the two sides. In general the direction of the polarization at any point may be orientated in any manner with reference to the normal to the surfaces, but the only case which is of any importance is that in which the resultant polarization is always directed along the normal to the shell at every point. Such shells are said to be normally polarized, and we shall confine our attention to them. They are of great importance in the theory of electromagnetism.

Let AD, BC be a section of the surfaces bounding the normally polarized shell, AD being positively and BC negatively charged. Let P be a distant point, OP being equal to r. O is any point in the substance of the shell and ON is the normal. AD, QR, AB, DC, etc. are infinitesimal. The angle

$$PON=\theta.$$

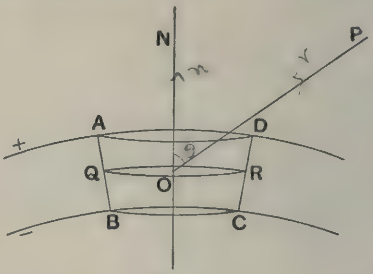


Fig. 15.

Let t be the thickness and P the polarization of the shell at O. Denote the element of area $AD=QR=BC$ by dS. Then the element of the shell $ADCB$ is equivalent to a doublet whose moment is $PtdS$. In dealing with shells it is convenient to introduce a new quantity called the strength of the shell.

The strength of a shell at any point is equal to the product of the intensity of the polarization of the shell by the thickness of the shell at that point.

We shall denote it by ϕ. Then $\phi=Pt$. Since $PtdS$ is the moment of a portion of the shell whose area is dS, the strength ϕ is equal to the moment of the shell per unit area.

Now consider the potential at P arising from the shell. We have seen (p. 58) that the potential at a point distant r

due to a doublet of moment μ is $\frac{\mu}{4\pi}\frac{\partial}{\partial s}\left(\frac{1}{r}\right)$, where $\frac{\partial}{\partial s}$ denotes differentiation along the axis of the doublet. In the case of the element of the shell $\mu = \phi dS$ and $\frac{\partial}{\partial s} = \frac{\partial}{\partial n}$, where $\frac{\partial}{\partial n}$ denotes differentiation along the direction of the outward normal to the positive face of the shell. Thus the potential due to the element dS is

$$dV = \frac{\phi}{4\pi}\frac{\partial}{\partial n}\left(\frac{1}{r}\right) dS,$$

and that due to the whole shell is

$$V = \iint \frac{\phi}{4\pi}\frac{\partial}{\partial n}\left(\frac{1}{r}\right) dS \quad\text{.....................(8)},$$

where the surface integral is extended over the whole of the positive surface of the shell.

The most important case which arises is that in which the strength ϕ has the same value at every point of the shell. The shell is then said to be uniform or of uniform strength. In such a case ϕ may be taken outside the integral, and

$$V = \frac{\phi}{4\pi}\iint \frac{\partial}{\partial n}\left(\frac{1}{r}\right) dS = \frac{\phi}{4\pi}\iint \frac{\cos\theta}{r^2}\, dS$$

$$= \frac{\phi}{4\pi}\iint d\omega = \frac{\phi\omega}{4\pi} \quad\text{..................................(9)},$$

where ω is the solid angle subtended by the entire shell at the point P.

We shall next calculate the potential energy of the shell in the field. First consider the potential energy of a doublet OQ which carries a charge $+e$ at Q and $-e$ at O. Let V_O, V_Q be the potentials at O and Q respectively. Then the potential energy of the doublet is

$$eV_Q - eV_O = e(V_Q - V_O) = e\,.\,OQ\frac{\partial V}{\partial s} = \mu\frac{\partial V}{\partial s}$$

$$= \mu\left(l\frac{\partial V}{\partial x} + m\frac{\partial V}{\partial y} + n\frac{\partial V}{\partial z}\right) \quad\text{............(10)},$$

where μ is the moment of the doublet and l, m, n the direction cosines of its axis s. Now apply this result to the case of the polarized shell. Considering the element of area dS of the shell $\mu \frac{\partial V}{\partial s} = \phi dS \times \frac{\partial V}{\partial n}$, so that the potential energy of the whole shell

$$= \iint \phi \frac{\partial V}{\partial n} dS \quad \text{......................(11).}$$

Since $-\frac{\partial V}{\partial n}$ is the force outward along the normal from the positive side of the shell, the surface integral represents the number of lines of force which thread the surface from the positive to the negative side. It is thus equal to $-\phi N$, where N is the number of tubes of force which leave the shell by the positive side.

Polarization on the Electron Theory.

The electron theory furnishes a very natural explanation of polarization. The chemical atoms out of which matter is built are regarded as consisting of a large number of electrified particles. The behaviour of these particles is considered to be quite different according to whether the substance is a conductor or an insulator of electricity. In conductors, part, at any rate, of the electrons are so loosely held that the very smallest electric field is sufficient to cause them to move about in the substance from one atom to another. In fact, as we shall see later, it is extremely probable that in conductors many of the electrons are always moving about inside, much in the same way as the molecules of a gas are believed to be in a state of continuous motion. The effect of an external field is simply to superpose on this haphazard motion an average flow in the direction of the field. This flow is what constitutes the electric current.

When an electric field is applied to an insulating substance the phenomena are different. The constituent electrons must, of course, be affected by the electric field, but none of them are able to move from one atom of the substance to another. In the absence of an electric field we regard the electrons as distributed about the atoms in positions of stable equilibrium. Under these circumstances the material exhibits no electric polarity. When an electric field is applied, the ultimate positive charges are pulled in the direction of the field, and the negative charges in the opposite direction. The displacement of the charges, however, is small, for they are pulled back by forces of the same nature as those which held them in equilibrium before the external field was applied. In the position of equilibrium which finally results, the force exerted on the electron by the external field will just balance the force tending to restore it to its original position of equilibrium.

It is clear from what has been said that the displacement of the ultimate electrified particles, which occurs in a dielectric when it is exposed to the action of an electric field, is equivalent to the creation of so many doublets, one for each particle. We have seen that the polarization which occurs in the dielectric under the same circumstances can be represented as due to the development of doublets in each element of volume. We shall now consider the whole matter from a more quantitative standpoint; as a result of our investigation we shall see that the results of the polarization theory can be obtained just as well from the properties of the doublets which develop from the displacement of the electrons.

Actual and Mean Values.

We have already pointed out (p. 9) that in the electron theory we have to deal with different elements of electric charge in different classes of problems which arise. A somewhat similar distinction arises in connection with many other physical quantities which determine the nature of the electric field. For instance in the discussion of this and the preceding chapter we have regarded the induction, the polarization and the electric intensity as vectors whose magnitudes changed only very gradually as we moved from one part of the field to another. We have always thought of them as though any alterations in their magnitudes which might occur, in a distance comparable with the distance between two molecules, could safely be considered as negligible, provided the two points compared were both in the same medium. This method of treatment obviously becomes inadequate when our view of the phenomena is so highly magnified as to take into account the effects of individual electrons or even atoms. So far, we have considered the space between two parallel planes filled with dielectric, when the planes are maintained at different potentials, as a region in which the electric intensity has the same magnitude and direction at every point. It is clear, however, that the actual electric intensity, the force exerted on a unit charge occupying an infinitesimal volume, will constantly change in both magnitude and direction from one part of this space to another. At places which are sufficiently close to an electron the actual force will be enormous compared with what we have called the electric intensity, and it may have any direction whatever.

One is tempted to ask what can be the use of a conception of the electric intensity which is so much at variance with what we believe to be the reality. The answer is, of course, that most of our methods of experimenting are so coarse, compared with the atomic scale, that they do not detect these enormous differences which occur within distances of the order of atomic magnitudes. Our experimental arrangements for the most part measure only the average values over spaces containing a large number of atoms. The reason why our average values possess validity is not because they are the true values but because, so far as such experimental arrangements enable us to detect, everything happens as if the average values were the true values.

It remains to specify the average values we have been dealing with more accurately than we have done hitherto. Let ϕ represent one of the scalar functions or a component of one of the vectors, which determine the state of the electric field. For example, ϕ may be the electrostatic potential at a point. Let τ be any small volume so chosen that its linear dimensions are large compared with atomic distances but small compared with the distances within which changes in ϕ are perceptible by the usual experimental methods. Then the average value of ϕ may be defined as the value of

$$\bar{\phi} = \frac{1}{\tau}\iiint \phi \, d\tau \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12),$$

where the integral is taken throughout the small volume τ. We evidently have

$$\overline{\frac{\partial \phi}{\partial \xi}} = \frac{1}{\tau}\iiint \frac{\partial \phi}{\partial \xi}\, d\tau = \frac{1}{\tau}\frac{\partial}{\partial \xi}\iiint \phi \, d\tau = \frac{\partial \bar{\phi}}{\partial \xi} \quad \ldots\ldots\ldots(13),$$

where ξ is any independent variable such as time, distance, etc. of which ϕ may be a function.

Since the actual potential V satisfies the relation $\nabla^2 V = -\rho$, it follows that $\nabla^2 \bar{V} = -\bar{\rho}$; and since $E = -\text{grad } V$, $\bar{E} = -\text{grad } \bar{V}$. Thus the average forces and potentials are the same as those which would obtain if the actual charges were replaced by a distribution of density equal to the average density at every point.

It is clear that the induction, polarization and electric intensity in a dielectric are average values in the sense indicated, and that the results that we have deduced are valid if this is understood.

Potential due to the Displaced Electrons.

We have seen that in the presence of an electric field the electrons are displaced, the positive in the direction of the field and the negative in the opposite direction. We shall see that the displacement thus produced is equivalent, for each electron, to the creation of a doublet of moment $\mu = es$, where e is the charge and s the displacement, of that electron. This doublet will contribute to the potential at a distant point P an amount $\frac{\mu}{4\pi}\frac{\partial}{\partial s}\left(\frac{1}{r}\right)$, and if there are ν such doublets per unit volume the total potential to which they will give rise at the point P will be

$$\frac{1}{4\pi}\iiint \nu\mu \frac{\partial}{\partial s}\left(\frac{1}{r}\right) d\tau.$$

In general the different electrons in the atom will be variously situated so that they will not all undergo the same displacement s in a given electric field. We may divide them up into classes, all the electrons in a class being characterised by a given value of s for a given field. Suppose there are n such classes and let ν_p, μ_p and s_p denote the values of ν, μ, s for the electrons of the pth class. Then it is clear that

$$4\pi V_P = \sum_{p=1}^{p=n}\iiint \nu_p\mu_p \frac{\partial}{\partial s_p}\left(\frac{1}{r}\right) d\tau \;\;\ldots\ldots\ldots\ldots\ldots(14).$$

We shall now consider the relation between the moment μ_p of the doublets and the electric field which produces them. The exact form of the relation between the restoring force and the displacement will depend on the arrangement of the electrons in the atom. At present our knowledge of this arrangement is very limited but, in any event, the restoring force must be a function of the displacement, which vanishes when the displacement is zero and is opposite in sign to the displacement to which it corresponds.

It follows from Taylor's Theorem that for small displacements the restoring force must be proportional to some odd power of the displacement; and since the frequency of the natural vibrations of bodies, as exemplified by their optical properties, is independent of the amplitude, it is natural to suppose that this power is the first. We shall assume, therefore, that when an electron of the pth class is displaced a small distance s_p the restoring force is equal to $+\frac{1}{\lambda_p} s_p$, where λ_p is a positive quantity which is constant and characteristic of this class of electrons. When the state of equilibrium is attained, the pull of the external electric field on the electron is balanced by the restoring force. The x component of the force on the electron is $-e_p \frac{\partial V}{\partial x_p}$, where V is the actual part of the potential at the electron whose charge is e_p which arises from the presence of the external field. If x_p is the x component of s_p then, provided the reaction to the displacement is independent of its direction in space, the x component of the restoring force measured in the positive direction of x is $-\frac{1}{\lambda_p} x_p$ and the equilibrium value of this displacement is

$$x_p = -\lambda_p e_p \frac{\partial V}{\partial x_p} \quad \text{.....................(15).}$$

Now a moment's consideration shows that when a charge e_p is displaced a distance x_p the electrical effect is exactly the same as that which is produced by the *creation* of a doublet whose moment is $e_p x_p$. For the displaced system is absolutely identical with that which is obtained when a doublet consisting of charges $\pm e_p$ at a distance x_p apart is superposed on the original system, in such a way that x_p coincides with the displacement x_p and the charge $-e_p$ coincides with the original charge $+e_p$. Thus the displacement (15) is equivalent to the creation of a doublet whose moment is

$$e_p x_p = -\lambda_p e_p{}^2 \frac{\partial V}{\partial x_p} \quad \text{.....................(16).}$$

In the absence of an electric field the medium is unpolarized and the potential due to the distribution of charges forming the atomic systems is zero. Thus the potential V_p due to the polarized medium is that which arises from the totality of the doublets which correspond to (16).

From equation (5) for the potential due to a single doublet we see that δV_P, the part of V_P which arises from the doublet the x component of whose moment is given by (16), is

$$\begin{aligned}\delta V_P &= \frac{e_p}{4\pi}\left\{x_p \frac{\partial}{\partial x_p}\left(\frac{1}{r}\right) + y_p \frac{\partial}{\partial y_p}\left(\frac{1}{r}\right) + z_p \frac{\partial}{\partial z_p}\left(\frac{1}{r}\right)\right\} \\ &= -\frac{\lambda_p e_p^2}{4\pi}\left\{\frac{\partial V}{\partial x_p}\frac{\partial}{\partial x_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial y_p}\frac{\partial}{\partial y_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial z_p}\frac{\partial}{\partial z_p}\left(\frac{1}{r}\right)\right\}.\end{aligned}$$

Now in addition to being made up of electrons the matter with which we are dealing possesses a coarser type of structure which we may term molecular. Each unit of molecular structure, which we may refer to as a molecule without necessarily thereby identifying it with the more definite chemical molecule, is characterized by the fact that, referred to its axes of symmetry, similarly situated electrons satisfy identical structural conditions. In considering a structure of this kind it is clearly absurd to endow the electrons, as we have done, with the property of suffering a hypothetical restoring force which, for a given displacement, is independent of the direction of that displacement in space. As we are confining ourselves to the case of non-crystalline substances this difficulty may be overcome by taking $s_p = x_p, y_p, z_p$ to be the average displacement of all the electrons which belong to the class p in a given small region and recollecting that all directions for the axes of symmetry of the molecules are equally probable. Suppose that each molecule contains n electrons, so that p has all the values from 1 to n, then ΔV_P the average contribution to V_P which arises from a molecule at the point a, b, c is

$$\Delta V_P = -\frac{1}{4\pi}\sum_{p=1}^{n}\lambda_p e_p^2\left\{\frac{\partial V}{\partial a_p}\frac{\partial}{\partial a_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial b_p}\frac{\partial}{\partial b_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial c_p}\frac{\partial}{\partial c_p}\left(\frac{1}{r}\right)\right\}.$$

If the dimensions of the element of volume $d\tau$ are small compared with $r = \{(x-a)^2 + (y-b)^2 + (z-c)^2\}^{\frac{1}{2}}$, where x, y, z are the coordinates of P and if it, nevertheless, contains a large number of molecules, the part dV_P of V_P which arises from the element of volume $d\tau$ is clearly

$$\begin{aligned}dV_P &= \nu \Delta V_P \, d\tau \\ &= -\frac{\nu}{4\pi}\sum_{p=1}^{n}\lambda_p e_p^2\left\{\frac{\partial V}{\partial a_p}\frac{\partial}{\partial a_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial b_p}\frac{\partial}{\partial b_p}\left(\frac{1}{r}\right) + \frac{\partial V}{\partial c_p}\frac{\partial}{\partial c_p}\left(\frac{1}{r}\right)\right\} d\tau,\end{aligned}$$

where ν is the number of molecules in unit volume. Thus

$$V_p = -\frac{\nu}{4\pi}\sum_{p=1}^{n}\lambda_p e_p{}^2\iiint\left\{\frac{\partial V}{\partial a_p}\frac{\partial}{\partial a_p}\left(\frac{1}{r}\right)+\frac{\partial V}{\partial b_p}\frac{\partial}{\partial b_p}\left(\frac{1}{r}\right)+\frac{\partial V}{\partial c_p}\frac{\partial}{\partial c_p}\left(\frac{1}{r}\right)\right\}da\,db\,dc \quad \text{......(17).}$$

By comparison with equations (7) and (16) we find that the components of the polarization P are

$$\left.\begin{aligned} P_x &= -\nu\sum_{p=1}^{n}\lambda_p e_p{}^2\frac{\partial V}{\partial a_p} = \nu\sum_{p=1}^{n} e_p x_p \\ P_y &= -\nu\sum_{p=1}^{n}\lambda_p e_p{}^2\frac{\partial V}{\partial b_p} = \nu\sum_{p=1}^{n} e_p y_p \\ P_z &= -\nu\sum_{p=1}^{n}\lambda_p e_p{}^2\frac{\partial V}{\partial c_p} = \nu\sum_{p=1}^{n} e_p z_p \end{aligned}\right\} \quad \text{............(18).}$$

The polarization is thus equal to the sum of the moments of all the equivalent doublets in unit volume. The dielectric coefficient κ is given by the relation

$$\left.\begin{aligned} \nu\sum_{p=1}^{n}\lambda_p e_p{}^2\frac{\partial V}{\partial a_p} &= -P_a = (\kappa-1)\frac{\partial \overline{V}}{\partial a} \\ \text{Thus}\qquad \kappa &= 1+\frac{\nu}{\dfrac{\partial \overline{V}}{\partial a}}\sum_{p=1}^{n}\lambda_p e_p{}^2\frac{\partial V}{\partial a_p} \end{aligned}\right\} \quad \text{.........(19).}$$

For crystalline media, λ_p will take different values for the different directions a, b and c because the axes of the molecules are definitely orientated in such substances.

Now if we average over a large number of molecules, $\dfrac{\partial V}{\partial a_p}$ under the sign of summation in (19) does not become equal to the value of $\dfrac{\partial \overline{V}}{\partial a}$ in the same region. There are two reasons for this. The first of these depends upon the definiteness of the arrangement of the electrons in the molecule. The electron whose type we have indicated by the suffix p is always subject, owing to the definite structure of the molecule, to certain geometrical relationships with the other electrons in the same molecule. This fact is not taken account of in the definition of $\overline{V}$ and of $\dfrac{\partial \overline{V}}{\partial a}$. The second reason is independent of the arrangement of the electrons within the molecule and is caused by the molecular rather than by the electronic structure of matter. The nature of this second factor can be most readily brought out by considering the dielectric properties of an ideal kind of matter whose imaginary molecules are so simply constituted that the first factor does not occur. We shall therefore consider the case of a substance whose molecules are monatomic and whose atoms give rise only to a single doublet each, under the influence of the electric field.

Case of the Ideal Simple Substance.

We suppose each molecule of the substance to consist of a single atom, and that, under the influence of the external field, each atom develops a single doublet placed at its centre. The forces acting on one of the electrons whose displacement gives rise to this doublet consist of

(1) the restoring force called into play by its displacement,

(2) the force arising from the charges in the field, including the doublets of the polarized medium not situated in its immediate neighbourhood,

(3) the force arising from the doublets in the immediate neighbourhood of the atom.

When equilibrium is established

$$(1) = (2) + (3).$$

It is clear from our discussion of the electric intensity in dielectric media that (2) is equal to the electric intensity $\bar{E} = -\dfrac{\partial \bar{V}}{\partial s}$. It remains to discover the nature of (3). About the doublet under consideration as centre describe a sphere whose radius is so small that the value of $\dfrac{\partial \bar{V}}{\partial s}$ does not vary appreciably in a distance comparable with it. At the same time the sphere must be big enough to contain a large number of molecules. The force (3) will be equal to the force exerted by the doublets in this sphere on the electron under consideration. This will only be true provided the dimensions of the sphere are within the assigned limits; otherwise this force will not be independent of the radius of the sphere.

To calculate the magnitude of (3) we suppose the spherical portion of the medium to be removed. The doublet now lies at the centre of a small spherical cavity. On account of this, and owing to the fact that the doublets behave on the average like the equivalent polarization P, the doublet at the centre will be acted on by a force, additional to $\bar{E}$, whose amount is determined by the equivalent polarization charge on the walls of the cavity. This is equal to $-P\cos\theta$ per unit area at any point, on the surface of the sphere, the radius to which makes an angle θ with the direction of P. The resultant force due to the whole distribution over the spherical surface is thus $\iint \frac{eP}{4\pi r^2}\cos^2\theta\, dS$, where the integral extends over the surface of the sphere whose radius is r. This is equal to $\frac{1}{3}Pe$ and is independent of the radius of the sphere.

The remaining part of (3) consists of the force which would be caused by the doublets which we have removed, if they had not been removed. This will depend very much on the geometrical arrangement of the atoms among one another. In certain particular cases this force vanishes. A doublet situated at a point whose coordinates are x, y, z, with respect to the centre as origin, and whose moment has components equal to μ_x, μ_y, μ_z, will give rise to a force at the centre, whose x component is

$$\frac{\mu_x}{4\pi}\frac{3x^2-r^2}{r^5}+\frac{\mu_y}{4\pi}\frac{3xy}{r^5}+\frac{\mu_z}{4\pi}\frac{3xz}{r^5},$$

where $$r^2=x^2+y^2+z^2.$$

If the atoms are arranged fortuitously so that any one position in the sphere is as likely as another the mean values $\overline{xy}=\overline{xz}=0$ and

$$\overline{\frac{3x^2-r^2}{r^5}}=\overline{\frac{3y^2-r^2}{r^5}}=\overline{\frac{3z^2-r^2}{r^5}}$$

$$=\overline{\frac{3\,(x^2+y^2+z^2)-3r^2}{3r^5}}=0.$$

It follows that the force arising from the doublets which we have removed out of the cavity vanishes on the average, if the atoms are arranged fortuitously. The same is true if they are arranged in regular cubical order*. It follows, in either case, if the atoms have the simple constitution we have imagined, that the force (3) $= \frac{1}{3} Pe$. When the molecules have a less symmetrical distribution, the additional force arising from the molecules which we took out of the spherical cavity will still be proportional to the polarization P, so that we can represent the more complex cases if we replace the factor $\frac{1}{3}$ by an unknown factor α depending on the configuration of the molecules.

---

* H. A. Lorentz, *Theory of Electrons*, p. 306.

Returning to the more symmetrical distribution, we see that the total force acting on the electron at the centre of the atom under consideration is, on the average,

$$-e \overline{\frac{\partial V}{\partial s_p}} = (E + \tfrac{1}{3} P)\, e \quad \text{.....................(20)},$$

so that comparing with formulae (16), (18) and (19), since $n = 1$,

$$\nu \lambda e^2 (E + \tfrac{1}{3} P) = P = (\kappa - 1) E$$

and
$$P = \frac{\nu \lambda e^2 E}{1 - \frac{1}{3} \nu \lambda e^2} \quad \text{.....................(21)},$$

whence
$$\frac{\kappa - 1}{\kappa + 2} = \frac{\lambda e^2 \nu}{3} \quad \text{...........................(22)}.$$

If we apply this formula to the case of a gas, we see that the only one of the quantities on the right hand side which varies with the density of the gas is ν, the number of molecules per cubic centimetre. This is proportional to the density, so that for a gas $\frac{\kappa - 1}{\kappa + 2}$ should be proportional to the density. The results of experiments are in agreement with this formula within the limits of experimental error, although the experimental measurements of the dielectric constants of gases are not very exact.

When we come to consider the phenomena of refraction and dispersion of light, we shall see that a very similar formula, in which κ is replaced by n^2, connecting the refractive index n with the density, can be developed along similar lines. It seems advisable to postpone the detailed discussion of the experimental evidence for and against these formulae until the optical phenomena are considered, as the evidence will then be much more complete.

We shall now return to consider the first of the two reasons why the average value of $\frac{\partial V}{\partial a_p}$ is not equal to $\frac{\partial \overline{V}}{\partial a}$. This one depends on the complexity of the atomic structure itself and not on the mere fact that matter possesses an atomic constitution. The nature of this factor can best be realized by considering a very exaggerated case. Suppose an atom to contain a very large number of electrons all very loosely held. If such an atom is placed in an electric field it will behave like a conductor of the same size and shape; so that there will be no field acting on the electrons in the interior. The electrons towards the outside of the atom will move so as to shield those inside from the action of the external electric force. The same effect will also occur to a smaller extent even when the number of electrons is comparatively small and their displacements are inconsiderable. It is clear that the average value of the force throughout a small volume of the material is different from the average value taken over a particular type of electron.

The force acting on an electron inside a molecule will arise partly from the charges outside the molecule and partly from the doublets inside the molecule itself. We can regard each molecule as equivalent to a simple atom possessing the same average electric moment, so that the force acting on an electron inside a molecule arising from external causes will be $e(E + \alpha P)$: where the constant α will depend on the geometrical configuration of the molecules. taking the value $\frac{1}{3}$ when the distribution is a fortuitous one as in a fluid. The way in which the second part of the force on the internal electron depends upon the external field may be realized by considering the conditions which are necessary in order to change the displacements of all the electrons in a given ratio. The displacements are proportional to the forces acting, so that this means that the force acting on an electron in the field will be changed in the same ratio at every point. Now the force arising from a given doublet is proportional to the moment of that doublet, so that the part of the force acting on the given electron which arises from other doublets in the same atom will be altered in the same ratio as the total force at any point in the field. It follows that the difference between this and the total force, which is the part of the force which is of external origin, must be changed in the same ratio. It follows from these considerations that however the external field may change, the force acting on any assigned electron will always be changed in the same proportion. This result may be represented by putting

$$-e_p \frac{\partial V}{\partial s_p} = L_p e_p (E + \alpha P) \quad \text{..................(23)},$$

where L_p is a constant characteristic of the pth class of electrons.

Comparing with p. 73 we see that

$$\nu \sum_{p=1}^{p=n} \lambda_p L_p e_p^2 (E + \alpha P) = P = (\kappa - 1) E,$$

so that

$$P = \frac{\nu \Sigma \lambda_p L_p e_p^2}{1 - \alpha \nu \Sigma \lambda_p L_p e_p^2} E \quad \text{..................(24)}$$

and

$$\begin{aligned} \frac{\kappa - 1}{\kappa + \dfrac{1-\alpha}{\alpha}} &= \frac{\Sigma \nu \lambda_p L_p e_p^2}{1 - \alpha \Sigma \nu \lambda_p L_p e_p^2} \\ &\quad \div \left(\frac{\Sigma}{1 - \alpha \Sigma} + \frac{1}{\alpha} \left(\frac{1 - \alpha \Sigma}{1 - \alpha \Sigma} \right) \right) \\ &= \alpha \nu \sum_{p=1}^{p=n} \lambda_p L_p e_p^2 \quad \text{....................(25)}. \end{aligned}$$

λ_p, L_p and e_p may vary for different electrons in the same molecule but they will have the same value for corresponding electrons in different molecules of the same substance. The expression on the right hand side may therefore be represented by a summation over each molecule multiplied by the number of molecules of the substance in unit volume. We therefore find that

$$\frac{\kappa - 1}{\kappa + \dfrac{1-\alpha}{\alpha}} = k\rho \quad \text{......................(26)},$$

where k is a constant and ρ is the density of the substance.

We shall find that the coefficients λ_p have an important significance when we come to consider the phenomenon of optical dispersion. The investigation leading up to formula (25) will not apply to optical problems without modification, as the displacement of the electrons in such cases is not necessarily always in phase with the corresponding "force."

When we are dealing merely with the electrostatic behaviour of dielectrics we can afford to neglect the complications just alluded to, which arise from the mode of arrangement of the electrons in the molecules. It is clear from the considerations which have been brought forward that instead of (16) we might have put

$$e_p x_p = -\sigma_p e_p^2 \left(\frac{\partial \overline{V}}{\partial x} - \alpha P_x\right) \quad \ldots\ldots\ldots\ldots\ldots(27),$$

where σ_p is a new constant, and thus obtained

$$\frac{\kappa - 1}{\kappa + \dfrac{1-\alpha}{\alpha}} = \alpha\nu \sum_{p=1}^{n} \sigma_p e_p^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(28),$$

so that

$$\sum_{p=1}^{n} \sigma_p e_p^2 = \sum_{p=1}^{n} \lambda_p L_p e_p^2.$$

However, the relation between σ_p and the natural frequencies of the material is not an obvious one without further investigation. This point is of importance in dealing with the corresponding optical phenomena.

CHAPTER V

MAGNETISM

We are not yet in a position to enter into the theory of magnetism in any fundamental way. We shall have to defer that until we have considered the phenomena which are grouped under the head of electromagnetism. It seems desirable, however, at this stage to enumerate some of the elementary facts and principles of magnetostatics, and to consider to what extent they resemble or differ from the corresponding phenomena of electrostatics.

One of the most striking of magnetic phenomena, and the first to be discovered, is the occurrence of intrinsic magnetization. A body exhibiting this phenomenon, usually called a permanent magnet, possesses polar properties. When two magnets are compared it is found that the ends A, A' of the first always exert forces of a certain kind on the ends B, B' of the second. Thus if A and B repel one another so do A' and B', whereas A' attracts B and A attracts B'. These relations between the two magnets are in fact the same as those between two similar portions of matter endowed with electric polarization.

Although this formal resemblance between magnetization and electrification exists, it is to be borne in mind that there is no static reaction between a magnet and an electric charge.

It is clear that the behaviour of magnets enables us to speak of the magnetism at the one end as positive, and that at the other end as negative. And here, at the outset, we meet one of the most striking differences between magnetization and electrification. It is impossible to separate the magnetic charges from one another. Every magnet carries equal and opposite magnetic charges. There is no force acting between two small magnets which varies inversely as the square of their distance apart except gravitational attraction. Thus a magnet is a body which is intrinsically polarized.

Although polarization is a common phenomenon in electrostatics, and occurs in every dielectric under the action of an external electric field, intrinsic polarization is comparatively unimportant. It probably occurs in crystals; where it appears to be required to explain the production of electrified surfaces by fracture and the phenomena of pyroelectricity. It is possible, however, that intrinsic dielectric polarization is commoner than is generally supposed. For unless a substance is a good insulator, it will always cover itself with a distribution of electric charge which is just such as is required to annul the external action of any intrinsic polarization it may possess. In all but a few cases, therefore, the existence of such a property would be difficult to detect.

On account of the fact that we are unable to separate the opposite magnetic charges, the investigation of their mutual forces is not so simple, in theory, as the corresponding electrical problem. In a very elaborate investigation Gauss examined the interaction between two magnets and showed that the forces were such as would arise if each element of magnetic charge repelled each like, and attracted each unlike, element with a force proportional to the product of the charges on each element, and inversely as the squares of their distances apart. This result was established with considerable accuracy.

We are now in a position to define our unit of magnetism. Consistently with our definition of the unit of electric charge we shall define it as that charge which repels an equal and similar charge at unit distance from it with a force equal to $1/4\pi$ dynes. To avoid the difficulty which is created by the inseparability of the magnetic charges from the opposite poles we may suppose that those between which the forces are measured are at the ends of infinitely long uniformly polarized magnets. The other ends will then be so far away that they will exert no influence.

Magnetic intensity is defined, in an analogous way to electric intensity, as the force exerted on a unit pole at any point of the field. It is convenient to use the term magnetic force rather than magnetic intensity in order to avoid confusion with intensity of magnetization, which is the name usually given to the quantity in magnetism which corresponds with intensity of polarization in electrostatics. The magnetic potential at a point is the work done in bringing a unit pole from a point at an infinite distance up to the point in question.

Just as in electrostatics we have to deal with dielectric media, so in magnetism there are media which, without being themselves permanently magnetized, have the power of modifying the distribution of magnetic force in their neighbourhoods. There is therefore a vector called magnetic induction analogous to the electric induction D. We shall denote the magnetic induction by the symbol B and its components by B_x, B_y, B_z. The coefficient which corresponds to the dielectric coefficient κ we shall call the permeability and denote by μ.

The behaviour of different media towards magnetic force furnishes more variety than the corresponding electrostatic phenomena. For a few substances μ may have very large values, and usually these are the substances which are capable of being permanently magnetized. Since iron is the typical example of this class of substances they are often called ferromagnetic. For all other substances μ does not differ greatly from unity, on the electromagnetic system of measurement. It is found that, in addition to the substances for which μ is greater than unity, and whose behaviour is analogous to that of dielectrics, there is another large class for which μ is less than unity. The former are said to be paramagnetic and the latter diamagnetic. It is very probable that paramagnetism and diamagnetism arise from the operation of separate causes. We shall, however, postpone the consideration of the physical causes which underlie the varied magnetic behaviour of substances until a later chapter. At present we are concerned with the formal relationship between magnetism and electrostatics.

If $H = H_x, H_y, H_z$ is the magnetic force it follows from the foregoing considerations, combined with the results of preceding chapters, that the following propositions are true for magnetism:

(1) The force between two poles of strengths m, m' embedded in a medium of permeability μ at a distance r apart is

$$F = \frac{mm'}{4\pi\mu r^2} \qquad \text{.........................(1).}$$

(2) The field of magnetic intensity can be mapped out by means of lines of force whose equations are

$$\frac{dx}{H_x} = \frac{dy}{H_y} = \frac{dz}{H_z} = \frac{ds}{H} \text{ ....(2).}$$

(3) The magnetic potential Ω at any point is

$$\Omega = -\int (lH_x + mH_y + nH_z)\, ds = -\int (H_x dx + H_y dy + H_z dz) \text{...(3).}$$

(4) The equation satisfied by Ω in free space is

$$\nabla^2 \Omega = 0 \text{ ....(4).}$$

(5) If dS is an element of any closed surface not intersecting a magnetic medium, that is to say, one which lies entirely in a medium of permeability unity, and if l, m, n are the direction cosines of the normal to this element, then

$$\iint (lH_x + mH_y + nH_z)\, dS$$

taken over the whole of the closed surface is equal to zero. This is the magnetic analogue of Gauss's Theorem for the free aether, and is true since each magnetic substance contains equal and opposite magnetic charges. It follows that in free space the magnetic intensity H is a solenoidal vector.

(6) The energy of a system of magnets may be obtained, as in the case of the system of electric charges, by bringing up equal fractions of the final system, one at a time, from a state of infinite dissemination. As we do not wish to contemplate the existence of separate magnetic charges it is desirable to regard the disseminated elements as magnets and not charges. This introduces the intensity of magnetization (I_x, I_y, I_z) instead of the density of charges into the final expressions. In this way the energy of a system of magnets is found to be

$$W = \frac{1}{2} \iiint \left(I_x \frac{\partial \Omega}{\partial x} + I_y \frac{\partial \Omega}{\partial y} + I_z \frac{\partial \Omega}{\partial z} \right) dx\, dy\, dz \text{ ......(5),}$$

the integral being taken throughout the magnetized matter. It follows from this expression, by a calculation similar to that carried out in the similar electrostatic case, that the energy per unit volume of the field in the free aether is

$$\tfrac{1}{2} H^2 \text{ ....(6).}$$

The points involved here are discussed at length by Jeans, *Electricity and Magnetism*, pp. 384—388.

(7) The magnetic forces can be represented as arising from the following system of Maxwell stresses in the aether:

$$\left.\begin{aligned} q_{xx} &= \frac{1}{2}\left\{\left(\frac{\partial\Omega}{\partial x}\right)^2 - \left(\frac{\partial\Omega}{\partial y}\right)^2 - \left(\frac{\partial\Omega}{\partial z}\right)^2\right\} \\ q_{yy} &= \frac{1}{2}\left\{\left(\frac{\partial\Omega}{\partial y}\right)^2 - \left(\frac{\partial\Omega}{\partial z}\right)^2 - \left(\frac{\partial\Omega}{\partial x}\right)^2\right\} \\ q_{zz} &= \frac{1}{2}\left\{\left(\frac{\partial\Omega}{\partial z}\right)^2 - \left(\frac{\partial\Omega}{\partial x}\right)^2 - \left(\frac{\partial\Omega}{\partial y}\right)^2\right\} \\ q_{xy} &= q_{yx} = \frac{\partial\Omega}{\partial x}\,\frac{\partial\Omega}{\partial y} \\ q_{xz} &= q_{zx} = \frac{\partial\Omega}{\partial z}\,\frac{\partial\Omega}{\partial x} \\ q_{yz} &= q_{zy} = \frac{\partial\Omega}{\partial y}\,\frac{\partial\Omega}{\partial z} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(7).$$

These are equivalent to a tension $\frac{1}{2}H^2$ per unit area along the lines of force and an equal pressure at right angles.

(8) The potential at a point, at distance r, due to a small magnet at a, b, c whose moment is $\mu = \mu_x, \mu_y, \mu_z$ is

$$\Omega = \frac{\mu}{4\pi}\frac{\partial}{\partial s}\left(\frac{1}{r}\right) = \frac{1}{4\pi}\left\{\mu_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) + \mu_y \frac{\partial}{\partial b}\left(\frac{1}{r}\right) + \mu_z \frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} \ldots(8).$$

(9) If $I = I_x, I_y, I_z$ is the intensity of magnetization (not magnetic intensity) the potential to which it gives rise at a point at distance $r = \{(x-a)^2 + (y-b)^2 + (z-c)^2\}^{\frac{1}{2}}$ is

$$\Omega = \frac{1}{4\pi}\iiint\left\{I_x \frac{\partial}{\partial a}\left(\frac{1}{r}\right) + I_y \frac{\partial}{\partial b}\left(\frac{1}{r}\right) + I_z \frac{\partial}{\partial c}\left(\frac{1}{r}\right)\right\} da\,db\,dc, \ldots(9),$$

I being the value at the point a, b, c.

(10) The potential due to a uniformly magnetized shell at a point at which it subtends a solid angle ω is

$$\Omega = \frac{\phi\omega}{4\pi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10),$$

where ϕ is the strength, or magnetic moment per unit area, of the shell.

(11) The potential energy of the shell in the magnetic field is

$$W = \iint \phi \frac{\partial\Omega}{\partial n}\, dS \quad \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

taken over one surface of the shell.

(12) The behaviour of the magnetic induction calls for rather more detailed consideration. From the electronic standpoint it is undesirable to define the induction as a vector flowing out in tubes from the magnetic charges. The reason for this, as will be clearer later, is that the conception of a magnetic charge is more artificial than that of an electric charge. We are on safer ground if we define the magnetic induction as the force in a flat cavity of the kind considered in Chapter IV, since we know that, when magnetized substances are fractured, opposite polarities develop on the new surfaces. It is clear from the reasoning in Chapter IV that the two modes of defining the induction are consistent with each other. It would be unfortunate were it otherwise, as we have to make use of the conception of a magnetic pole as the foundation of all our magnetic measurements.

It is clear that the whole of the theorems which we have grouped under the heading of Poisson's Theory of polarized media may be transferred bodily to magnetism if we replace polarization by magnetization. In fact the theory was originally developed by Poisson as a theory of magnetostatics. It follows that when the magnetization is uniform there are surface charges but no volume charges, and that in any event the algebraic sum of all the charges of any distribution of magnetization is zero. Also if H, B and I are the magnetic force, the magnetic induction and the intensity of magnetization at any point of any medium,

$$B = H + I = \mu H \quad \text{..................(12)},$$

and

$$I = (\mu - 1) H \quad \text{......................(13)}.$$

These are vector equations and are each equivalent to three equations between the corresponding components of the vectors.

Let us now describe any closed surface S intersecting the magnetic media B, C, etc. in D, E, etc. Let us cut away an infinitesimal layer of each of the magnetic media at both sides of the intersection. Then the surface lies entirely in the medium whose permeability is unity, so that if F_x, F_y, F_z are the components of the force on a unit magnetic pole at any point of the surface

$$\iint (lF_x + mF_y + nF_z)\, dS = 0,$$

since the total magnetic charge inside the surface is zero (Proposition (5)). But since the cavities are cut parallel to the surface, F_x, F_y, F_z are *actually* the components of the induction B_x, B_y, B_z. It follows that over any closed surface in space

$$\iint (lB_x + mB_y + nB_z)\,dS = 0 \quad\text{..............(14).}$$

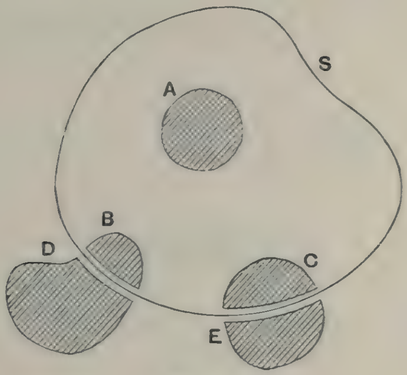


Fig. 16.

It follows that
$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \quad\text{....................(15)}$$

and the magnetic induction B is solenoidal *everywhere*. The corresponding result only holds for the electric induction in regions which do not contain any electric charges. Where there are "true" charges of volume density ρ

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho.$$

Force on a Magnetic Shell.

We have seen that the potential energy of a uniform magnetic shell of strength ϕ is $-N\phi$, where N is the number of lines of magnetic force threading it in the positive direction. This result enables us to calculate the force on the shell in the field, and we shall see that it may be represented as so much per unit length of the boundary of the shell. Let the shell be displaced so that N becomes $N + \delta N$. The diminution of the potential energy due to the displacement will be $\phi\delta N$, and this will be equal to the work done on the shell by the magnetic forces. The value of δN is easily calculated. Consider the prismatic figure, bounded by the original and displaced positions of the shell, which is traced out by the edge of the shell during the displacement. We shall suppose the displacement to be small and rectilinear; let us denote it by δx. Since there are no magnetic charges inside this figure the total flux of force over its boundary is zero. Hence

$$N + \delta N - N + \iint H \cos \widehat{nH} \sin \widehat{xs}\, ds\, \delta x = 0,$$

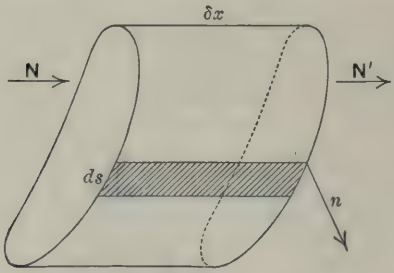


Fig. 17.

where H is the magnetic force at any point s of the boundary, $\widehat{xs}$ is the angle between the direction of the displacement δx and the element ds of the boundary, and $\widehat{nH}$ is the angle between the direction of H and the normal to the element of area $ds\,\delta x \sin \widehat{xs}$.

Thus $$\delta N = -\, \delta x \int H \cos \widehat{nH} \sin \widehat{xs}\, ds.$$

But if X is the force acting on the whole shell in the direction δx

$$X \delta x = \phi \delta N,$$

and $$X = -\, \phi \int H \cos \widehat{nH} \sin \widehat{xs}\, ds.$$

Thus the force X is equivalent to a set of forces of amount

$$-\, \phi H \cos \widehat{nH} \sin \widehat{xs}$$

per unit length of the edge of the shell. Since x is arbitrary this gives the force in any direction. To find the direction of the resultant force we notice that when x and s are in the same direction $\sin \widehat{xs} = 0$, so that the resultant force is at right angles to ds. It is also normal to H, since it vanishes when $\widehat{nH} = \pi/2$. It is therefore along the common normal to H and ds. To find the magnitude of this resultant let us suppose that the displacement δx is in the direction of this common normal. Then $\widehat{xs} = \pi/2$

and $\hat{nH} = \pi/2 - \hat{Hs}$, where $\hat{Hs}$ is the angle between H and ds. Thus the force on a length ds of the edge of the shell is

$$-\phi H \sin \hat{Hs} ds,$$

and is at right angles both to H and ds. In vector notation this result may be written more briefly as

$$-\phi [H . ds].$$

If the force H arises from a single magnetic charge m placed at a point P distant r from ds, then H acts along r and is equal to $\frac{m}{4\pi r^2}$. Thus the force exerted by m on an element ds of the boundary of the shell is equal to $-\frac{\phi m}{4\pi r^2} \sin \hat{rs} ds$. This must be equal and opposite to the force exerted by the element of the shell on m. Hence the force on the pole is equivalent to a series of forces of amount $\frac{\phi m}{4\pi r^2} \sin \hat{rs} ds$ arising from each element ds of the boundary of the shell. The magnetic intensity at P due to each element ds of the boundary is thus

$$\frac{\phi}{4\pi r^2} \sin \hat{rs} ds,$$

and is perpendicular to both r and ds.

These results have important applications in the theory of electromagnetism.

CHAPTER VI

ELECTROMAGNETISM

The Magnetic Potential due to an Electric Current.

In 1820 Oersted showed that an electric current gave rise to forces acting on the poles of a magnet placed in its neighbourhood. Thus an electric current gives rise to a distribution of magnetic intensity. Since the difference in the magnetic potential between two points is the work done in taking a unit pole from one point to the other, it follows that the magnetic intensity is the space derivative of the magnetic potential. The potential is essentially a scalar quantity. It is clear therefore that the magnetic potential at any point due to an electric current is a quantity which possesses magnitude but not direction, and which depends only on the position of the point relative to the circuit carrying the current and on the magnitude of the current. If we can determine the potential at every point of space arising from the electric current, we can deduce from it the distribution of magnetic intensity. We shall now consider how the magnetic potential may be calculated.

We shall base our demonstration on two empirical generalizations which are the result of experiment. They are :—

(1) That the magnetic force arising from an electric current flowing round a given circuit is proportional to the magnitude of the current.

(2) That any circuit carrying a given current can be replaced by one which carries the same current and continually zig-zags across it to a small distance on each side, without altering the magnetic force to which it gives rise.

The first generalization was established by the experiments of Faraday and the second by those of Ampère. It follows from (1) that two equal and opposite currents flowing in the same circuit or in the same branch of a circuit give rise to no magnetic force. This must be the case since the total current is then zero.

Let us take a large number of pairs of points on the boundary of any circuit and join each pair by a line. Suppose that along each of the lines there are two opposite currents flowing, each equal in magnitude to that flowing round the boundary. The lines we have drawn have divided the entire circuit into a series of small areas bounded by the lines. Round each small area there will be a current flowing equal in magnitude to the original current and circulating in the same sense about the normal to the circuit. It follows that the magnetic effect of any circuit is the same as that of any number of small circuits, bounded by the same contour, into which it may be completely subdivided, provided each circuit carries the same current as the original circuit, and that all the currents flow in the same sense about the normal to the imaginary surface in which they lie. The potential due to the whole circuit must therefore be equal to the sum of the potentials due to the constituent circuits.

We shall now show that the potential due to any circuit lying in a plane is zero at every point in its own plane. First consider the potential at the point O due to the circuit $ABCD$ (Fig. 18) bounded by arcs of circles whose centres are at O, and by radii passing through O. Since all the forces in the field are reversed when all the currents are reversed, it follows that the potential at every point will reverse when all the currents reverse. The potential at the point O will therefore reverse if the current $ABCD$ is reversed in direction. But the direction of the current can be reversed by simply rotating the circuit $ABCD$ through the angle π about the line of symmetry through O lying in the plane of the paper. But since potential has only magnitude and not direction, there is nothing which enables the potential at the point O, as it were, to tell which way up the circuit $ABCD$

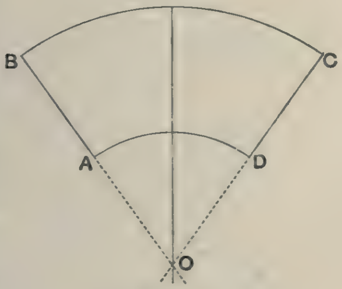


Fig. 18.

lies. The potential must therefore have the same value after it is reversed as before. Its value must therefore be zero.

Now if *ABCD* is replaced by a coplanar circuit of any shape, by drawing a large number of radii from *O* it can be replaced by equivalent circuits bounded by radii and its original contour. The contour can be replaced by a series of concentric arcs since the original circuit will give rise to the same field if it is caused to zig-zag about its original path. Each of these constituent circuits gives rise to zero potential at the point *O*, so that the whole circuit gives rise to zero potential at *O*.

Now consider any circuit *ABCD* (Fig. 19) lying in space. Describe the cone on which the circuit lies and whose apex is at *O*. Also describe a sphere of unit radius about *O*. Let this cut the cone in the curve *EFGH*. Let i be the current round *ABCD*. From *O* draw a series of lines *OEA*, *OFB*, etc. intersecting the curves in *E*, *F*, *A*, *B*, etc. Imagine equal and opposite currents i to flow along each of these lines. Suppose also that a current i flows around the curve *GFEH* in the direction from *F* to *E*. This is the only new current that has been added; since the equal and opposite currents along *FB*, etc. cancel each other. The system now contemplated resolves itself into a series of currents i round circuits such as *ABFE*. Since *O* lies in the plane of each of these circuits the potential at *O* due to the system of currents is zero. The potential at *O* due to i round $F \rightarrow EHG$ is thus equal and opposite to that round $A \rightarrow BCD$. The potential due to i round

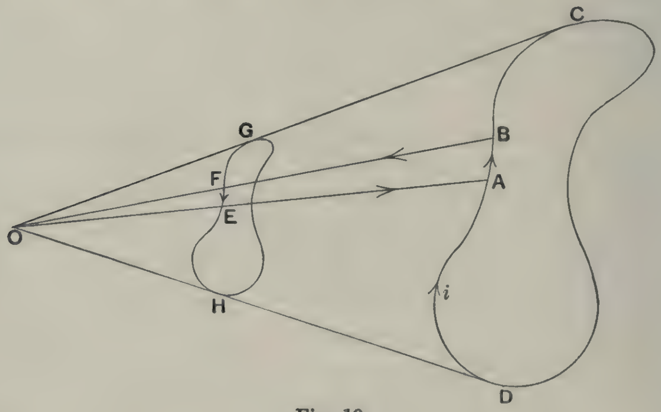


Fig. 19.

$A \rightarrow BCD$ is therefore the same as that due to the same current round $E \rightarrow FGH$. Now let us fill up the whole of the surface, cut off from the sphere of unit radius by the curve $EFGH$, with equal contiguous and superposable small areas. Imagine a current equal to i to flow round the boundary of each area. This will leave the original current i flowing round the boundary, and two equal and opposite currents along every line inside it. The potential due to the original circuit will therefore be equal to the potential due to the sum of the small circuits. Since all the small circuits are equal and have the same geometrical relation to O, the potential due to each one of them must be the same. It must be proportional to the current i since the force due to any system of currents changes in the same ratio as the currents when all the currents are changed in equal proportion. The argument is independent of the shape of the small circuits, and the only other condition to be satisfied is that the potential due to the whole circuit is the sum of those due to its parts. It follows that the geometrical factor to which the potential of each constituent circuit is proportional is $\delta\omega$, the solid angle it subtends at O. We therefore conclude that the potential due to each elementary circuit is $Ai\delta\omega$, where A is a constant which depends neither on the magnitude of the current nor on the geometry of the circuit.

It follows that when the current subtends a finite solid angle ω the potential to which it gives rise is $Ai\omega$. This is universally true since we have proved that all circuits carrying equal currents and lying on the same cone whose apex is at O, give rise to equal potentials at O. The magnitude of the constant A depends on the unit in which we measure the current. The usual electromagnetic unit of current is defined by making $A=1$. In the units used in this book $A=(4\pi c)^{-1}$, where c is the velocity of light. (See later, p. 112.)

In our study of the properties of polarized shells we saw that the potential of a shell of strength ϕ was equal to $\phi\omega/4\pi$, where ω was the solid angle subtended by the shell at the point where the potential was measured. An electric current is thus equivalent in its magnetic action to a magnetic shell whose strength is equal to $4\pi A$ times the intensity of the current.

There is an important difference between the field due to an electric current and that arising from the equivalent magnetic shell. In the latter case the work done in taking a unit pole round a closed curve which passes once through the shell is zero, since the work done against the forces outside the shell is just equal to the work done by the forces inside the shell. In the case of the electric current there is nothing which fixes the position of the equivalent shell in space except the current which forms its boundary. It follows that there cannot be any region in which the force has the peculiar distribution characteristic of the interior of an actual magnetic shell. The force due to a current will therefore be continuous everywhere, and the work done in going round any closed path which embraces the current once will be equal to $4\pi Ai$. If the path followed by the unit pole circulates round the current m times in the positive sense and n times in the negative sense, the work done on it will be $4\pi A(m-n)i$. Thus the magnetic potential due to a current is determined not solely by the relative position of the current and the point, but also by the number of times the path of the point has previously encircled the current circuit. It is what is known as a multiple-valued function of space. The complete expression for the magnetic potential at any point due to a current may be written

$$Ai\,[4\pi(m-n)+\omega].$$

The force due to a current does not, of course, depend on m or n but only on ω, so that it will be single-valued and will depend only on the intensity and geometrical distribution of the current and the position of the circuit relative to the point.

The foregoing result that the work done in taking a unit magnetic pole once round any closed path embracing a current is proportional to the current embraced is not confined to linear currents, but is true if the currents occupy a finite volume. This is obvious if we divide up the whole current i into linear constituents δi bounded by tubes of flow. Then the work done in taking a unit pole round a path enclosing one of the constituents will be $4\pi A\delta i$ and if the path encloses the whole current i it will be

$$4\pi A\,\Sigma\,\delta i = 4\pi Ai \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1).$$

The consideration of non-linear currents enables us to express this result in rather different analytical form. Consider any surface S in space, traversed by currents and bounded by a closed contour s. If the current is flowing in any specified direction at any point, the amount of electricity transported across unit area perpendicular to that direction in unit time is called the current density at that point. It is a vector quantity; let us denote it by j. Let $H = (H_x, H_y, H_z)$ be the magnetic intensity at any point. Then the work done in taking a unit magnetic pole around the boundary s will be

$$\int H \cos(\widehat{Hs})\, ds = \int \left\{ H_x \frac{dx}{ds} + H_y \frac{dy}{ds} + H_z \frac{dz}{ds} \right\} ds$$

$$= \int_s (H_x dx + H_y dy + H_z dz).$$

What is the total current i across the area S? If θ is the angle between the normal to an element dS of the surface, and j the resultant current density at that element, the current δi across the element is $j \times dS \cos\theta$. But $j\cos\theta = j_n$ the component of the current density normal to the element. The total current i through the entire area may therefore be written $i = \iint j_n dS$. Hence

$$\int (H_x dx + H_y dy + H_z dz) = 4\pi A \iint j_n dS \quad \text{.........(2).}$$

Before discussing this equation further we shall prove an important geometrical theorem, due to Stokes, connecting the line and surface integrals of vector point functions.

Stokes's Theorem.

Let R be any vector point function which is continuous throughout the region considered. Consider the value of the line integral $I = \int_P^Q R \cos \widehat{Rs}\, ds$ taken along any path PTQ from P to Q. $\widehat{Rs}$ is the angle between R and the tangent at a point of the path. Let us find the variation of the integral when the path of integration is changed by an infinitesimal amount, so as to lie along the curve PUQ. PUQ is only slightly displaced from PTQ, and the terminal points P and Q are not varied. Then

$$\begin{aligned}\delta I &= \delta\int_P^Q R\cos\hat{Rs}\,ds\\ &= \delta\int_P^Q\left(X\frac{\partial x}{\partial s}+Y\frac{\partial y}{\partial s}+Z\frac{\partial z}{\partial s}\right)ds\\ &= \int_P^Q\left\{X\frac{\partial}{\partial s}(\delta x)+\delta X\frac{\partial x}{\partial s}+Y\frac{\partial}{\partial s}(\delta y)+\delta Y\frac{\partial y}{\partial s}\right.\\ &\qquad\qquad\left.+Z\frac{\partial}{\partial s}(\delta z)+\delta Z\frac{\partial z}{\partial s}\right\}ds,\end{aligned}$$

if X, Y and Z are the components of R, since

$$\delta\left(\frac{\partial x}{\partial s}\right)=\frac{\partial}{\partial s}(\delta x),\quad \delta\left(\frac{\partial y}{\partial s}\right)=\frac{\partial}{\partial s}(\delta y)\ \text{ and }\ \delta\left(\frac{\partial z}{\partial s}\right)=\frac{\partial}{\partial s}(\delta z).$$

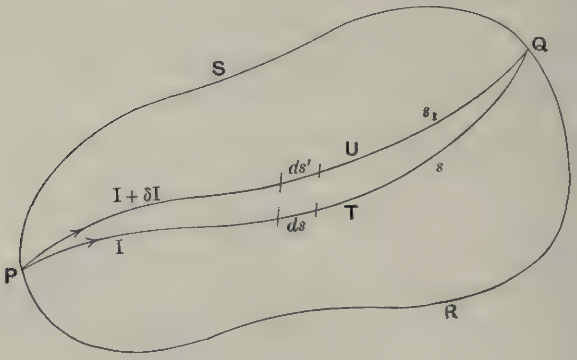


Fig. 20.

Let us integrate $\int_P^Q X\frac{\partial}{\partial s}(\delta x)\,ds$ by parts. The integrated part is $\Big|X\,\delta x\Big|_P^Q$ and the unintegrated part is $\int_P^Q\left(\delta x\frac{\partial X}{\partial s}\right)ds$.

Now $\frac{\partial X}{\partial s}=\frac{\partial X}{\partial x}\frac{\partial x}{\partial s}+\frac{\partial X}{\partial y}\frac{\partial y}{\partial s}+\frac{\partial X}{\partial z}\frac{\partial z}{\partial s}$, so that

$$\int_P^Q X\frac{\partial}{\partial s}(\delta x)\,ds=\Big|X\,\delta x\Big|_P^Q-\int_P^Q\delta x\left(\frac{\partial X}{\partial x}\frac{\partial x}{\partial s}+\frac{\partial X}{\partial y}\frac{\partial y}{\partial s}+\frac{\partial X}{\partial z}\frac{\partial z}{\partial s}\right)ds.$$

We obtain precisely similar expressions for

$$\int_P^Q Y\frac{\partial}{\partial s}(\delta y)\,ds\ \text{ and }\int_P^Q Z\frac{\partial}{\partial s}(\delta z)\,ds.$$

Now $\delta X = \frac{\partial X}{\partial x}\delta x + \frac{\partial X}{\partial y}\delta y + \frac{\partial X}{\partial z}\delta z$, so that

$$\delta X \frac{\partial x}{\partial s} = \frac{\partial X}{\partial x}\frac{\partial x}{\partial s}\delta x + \frac{\partial X}{\partial y}\frac{\partial x}{\partial s}\delta y + \frac{\partial X}{\partial z}\frac{\partial x}{\partial s}\delta z,$$

and similar expressions are obtained for $\delta Y \frac{\partial y}{\partial s}$ and $\delta Z \frac{\partial z}{\partial s}$. Collecting together all the terms which contain δx as a factor, and similarly those which contain δy and δz, we have

$$\begin{aligned}\delta I = &\left| X\delta x + Y\delta y + Z\delta z \right|_P^Q \\ &-\int_P^Q \left[\delta x\left\{\frac{\partial X}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial X}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial X}{\partial z}\frac{\partial z}{\partial s} - \frac{\partial X}{\partial x}\frac{\partial x}{\partial s} - \frac{\partial Y}{\partial x}\frac{\partial y}{\partial s} - \frac{\partial Z}{\partial x}\frac{\partial z}{\partial s}\right\}\right. \\ &\left. + \delta y\left\{\qquad\qquad\right\} + \delta z\left\{\qquad\qquad\right\}\right] ds.\end{aligned}$$

The terminal points P and Q are fixed, so that δx, δy and δz are zero at P and Q. Thus the integrated part vanishes, and arranging the terms rather differently we get:

$$\begin{aligned}\delta I = &\int_P^Q \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right)(\delta x\, dy - \delta y\, dx) \\ &+ \int_P^Q \left(\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}\right)(\delta y\, dz - \delta z\, dy) + \int_P^Q \left(\frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}\right)(\delta z\, dx - \delta x\, dz).\end{aligned}$$

To find the value of $\delta x\, dy - \delta y\, dx$ consider the parallelogram of which ds in its original and displaced positions forms two opposite sides. The angular points of this parallelogram are defined by s, $s + ds$, $s_1 (= s + \delta s)$ and $s_1 + ds_1 (= s + \delta s + ds)$. Let $ABDC$ (Fig. 21) be the projection of this parallelogram on the plane of xOy. AB corresponds to ds and CD to ds_1. Draw DK, BH parallel to Oy and BK, AH parallel to Ox, CF parallel to AH, CG perpendicular to AH, and produce DB to meet AH in E. Then

$$DK = \delta y,\ BK = \delta x,\ BH = dy \text{ and } AH = dx.$$

Also the parallelogram

$$\begin{aligned}ABDC &= AEFC = AE \times CG = CG \times (AH - EH) \\ &= \delta y\, dx - CG\,.\,EH = \delta y\, dx - BK\,.\,BH = \delta y\, dx - \delta x\, dy.\end{aligned}$$

Now if dS is the area of the original parallelogram, the parallelogram $ABDC = -dS \cos nz$, where nz is the angle between the normal to the original parallelogram dS and the axis of z, the normal being drawn downwards. So that

$$\delta y\,dx - \delta x\,dy = -\,dS\cos nz.$$

Similarly $\delta z\,dy - \delta y\,dz = -\,dS\cos nx$, and so on. Thus

$$\delta I = \int_P^Q \left\{\left(\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}\right)\cos nx + \left(\frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}\right)\cos ny + \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right)\cos nz\right\} dS.$$

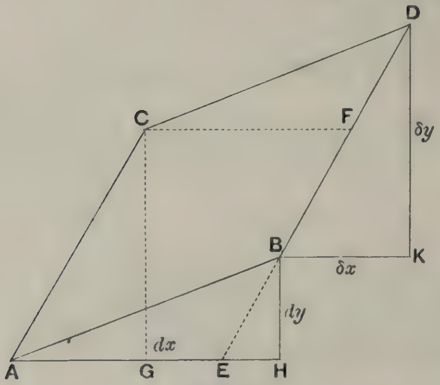


Fig. 21.

Next consider any closed contour $PRQS$ (Fig. 20) and let P and Q be any two points on it. From P to Q draw an indefinitely large number of paths such as PTQ, PUQ an infinitesimal distance apart. The difference between the value of the integral I taken along PSQ and along PRQ will be the sum of the differences along the infinitesimal paths. This is clear, since integration from P to Q and back again from Q to P along the same path adds nothing to the integral. The difference between the integral along PSQ and that along PRQ, starting at P and ending at Q in each case, is equal to the same integral taken all round the circuit in the direction PSQ; since the value of I from P to Q along PRQ is equal to $-I$ taken from Q through R to P.

Hence for the entire closed contour we have

$$\begin{aligned} I &= \int X\,dx + Y\,dy + Z\,dz \\ &= \Sigma\,\delta I \\ &= \iint\left\{\left(\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}\right)\cos nx + \left(\frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}\right)\cos ny + \left(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}\right)\cos nz\right\} dS \quad\text{...(3),}\end{aligned}$$

where the surface integral extends over any surface bounded by the contour, and where the cosines are the direction cosines of the normal to the surface at the point of integration. It will be observed that the positive direction of the normal is towards that side of the surface from which a right-handed screw would move if it were turned in the same sense as that of the integration round the contour.

The First Law of Electrodynamics.

Let us apply Stokes's Theorem to the case in which R is the magnetic intensity H whose components are H_x, H_y, H_z. We then have

$$\int (H_x dx + H_y dy + H_z dz)$$

$$= \iint \left\{ \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \right) \cos nx + \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right) \cos ny + \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right) \cos nz \right\} dS.$$

But we have seen that

$$\int H_x dx + H_y dy + H_z dz = 4\pi A \iint j \cos \widehat{nj}\, dS,$$

where j is the current density at any point of any surface bounded by the contour and $\widehat{nj}$ is the angle between the resultant current density and the normal to the surface at the point. Now

$$j \cos \widehat{nj} = j_x \cos nx + j_y \cos ny + j_z \cos nz,$$

where $\cos nx$, etc. are the direction cosines of the normal. It follows that

$$\left.\begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= 4\pi A j_x \\ \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= 4\pi A j_y \\ \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= 4\pi A j_z \end{aligned}\right\} \qquad \text{.....................(4)}$$

everywhere.

These differential equations are the most general expression of the distribution of the field of magnetic intensity due to a distribution of electric currents. They are one form of the first of Maxwell's two famous electromagnetic equations. To distinguish them from the second equation we shall sometimes refer to them as the statement of the First Law of Electrodynamics.

Like many of the formulae with which we have already become familiar, these three equations represent relations between components and derivatives of vectors. We shall frequently find it convenient to represent them by the abbreviated notation

$$\text{rot}\, H = 4\pi A j,$$

where by rot H we mean the vector whose components are the quantities on the left hand side of equations (4), taken in succession. This equation is a vector equation, that is to say, it holds independently for each of the components of rot H and j respectively.

The Electric Current.

The electron theory regards the electric current which flows along a wire as a convection current. It supposes that in a conductor there are a number of charged particles which do not execute small displacements about a position of equilibrium, as in the case of those electrons which we considered when we were discussing the behaviour of dielectric media, but which are able to move freely from one part of the conductor to another. In a metallic conductor these particles are believed to be electrons and are called "free electrons" to distinguish them from the bound electrons, which only undergo small displacements from their position of equilibrium when an electric field is made to act on them. The free electrons in a metal are believed to be in much the same condition as the molecules of a gas. When we come, later on, to consider the evidence for this belief, we shall see that it is very strong and that the resemblance is a very close one. In the case of liquid electrolytes the charged particles are of atomic or molecular dimensions, and in many cases of electric conduction through gases this is the case also.

In all these cases the charged particles, of whatever nature, are believed to be moving about irregularly in all directions, even when they are not subjected to the action of an electric field. This motion does not cause any transportation of electricity since on the average, the particles are just as likely to be moving in any one direction as in any other. The effect of an external electric field is to superpose on the irregular motion a definite drift, so that on the average the positively charged particles move in the direction of the electric field, and the negatively charged particles move in the opposite direction. In general there may be different kinds of ions of one sign present. Suppose that in the unit of volume there are N_p positive ions of type p, that their average velocity of drift under the applied field in the direction of the field is U_p, and that their charge is E_p. Let the corresponding quantities for the negatively charged carriers be denoted by small letters. The current density at any point will be the total amount of positive electricity transported across unit area perpendicular to the direction of the field at that point, in the positive direction, plus the total amount of negative electricity transported in the opposite direction. This is clearly

$$j = \Sigma N_p E_p U_p + \Sigma n_p e_p u_p \quad \dots\dots\dots\dots\dots\dots (5),$$

since U_p is the volume of a cylinder whose axis is parallel to the field, whose sectional area is unity and whose height is U_p, the velocity of drift of the particles under consideration. In the case of solid and liquid conductors the N's and E's are independent of the electric intensity whilst the U's are proportional to it, so that the current in these cases obeys Ohm's Law. The same is true in the case of very small currents in gases at moderate pressures. In general, however, in the case of gases both the N's and the U's may vary in a complicated way with the applied electromotive force. It is for this reason that the relation between the electromotive force and the current in gaseous conduction is, generally speaking, quite intricate.

The current whose properties we have been discussing is often called the true current. There is another kind of electric current called the displacement current, for the conception of which we are indebted to the constructive imagination of Maxwell.

The Displacement Current.

Consider an electric circuit consisting of a battery A, a condenser B and a one-way switch C. When the key C is depressed, a current flows through the wires from the battery into the plates of the condenser B. This current is not, however, continuous all the way round the circuit, in the same way that it would be if the condenser were replaced by a wire, since there is no actual transportation of electricity across the dielectric between the plates of the condenser. Maxwell put forward the hypothesis that the magnetic effect of such a circuit was identical with that arising when the condenser is replaced by a suitable conductor, and there is the same current flowing along the wires in the circuit. This comes about owing to the existence, when the field is changing, of what Maxwell called a displacement current in the dielectric. The displacement current density at any point is equal to $\frac{dD}{dt}$, where D is the electric induction.

This value for the displacement current makes the current round the circuit continuous. For if σ is the charge per unit area of the plates of the condenser at any point, the current i in the wires is equal to $\iint \frac{d\sigma}{dt}\, dS$, the integral being taken over a plate of the condenser. But, if D is the value of the induction, the displacement current close to the plates is $\iint \frac{dD}{dt}\, dS$ taken over the same surface. We have seen that $D = \sigma$, so that the displacement current is equal to i and the current is continuous all round the circuit.

This result may be proved to be quite general as follows:

Consider any closed surface in space. By Gauss's Theorem the charge e inside this surface is equal to

$$\iint (l D_x + m D_y + n D_z)\, dS.$$

The current flowing into the surface is

$$\frac{de}{dt} = \iint \left(l \frac{\partial D_x}{\partial t} + m \frac{\partial D_y}{\partial t} + n \frac{\partial D_z}{\partial t} \right) dS.$$

But if j_x, j_y, j_z are the components of the true current density

$$\frac{de}{dt} = - \iint (l j_x + m j_y + n j_z)\, dS.$$

Hence

$$\iint \left[l \left(j_x + \frac{\partial D_x}{\partial t} \right) + m \left(j_y + \frac{\partial D_y}{\partial t} \right) + n \left(j_z + \frac{\partial D_z}{\partial t} \right) \right] dS = 0 \quad \ldots(6).$$

So that the vector $i = j + \frac{\partial D}{\partial t}$ is solenoidal everywhere. Like the tubes of magnetic induction, the tubes of flow of the total current are always closed regions of space. They have no free ends, unless at infinity.

It would be very difficult to devise a direct experimental proof of the magnetic effect of displacement currents. The best proof of it is an inductive one. The hypothesis of displacement currents forms the basis of the electromagnetic theory of light, and the extraordinary way in which the requirements of this theory have been fulfilled by experiment shows that it is built upon a solid foundation.

It is often desirable to consider the displacement current density $\frac{dD}{dt}$ as made up of two parts, (1) the aethereal displacement current $\frac{dE}{dt}$ and (2) the polarization current $\frac{dP}{dt}$. Since $D = E + P$, $\frac{dD}{dt}$ is always equal to the sum of the aethereal current and the polarization current. On the electron theory, as is obvious from the discussion in Chapter V, the polarization current corresponds to an actual displacement of charged electrons, and is to that extent very similar to the true current.

Convection Currents.

The kinds of electric currents which flow in wires, electrolytic cells and so forth, and which are carried by extremely minute particles, are not the only ones whose magnetic effects can be detected. Rowland* showed in 1876 that an electrostatically charged disc when made to revolve at a sufficiently high speed affected a suspended magnet in the same way as a current flowing round the disc. Effects of this kind, that is to say, effects depending on the movement of electricity on a large scale, may easily be summarized by means of a very simple formula. Let ρ be the net volume density of the electrification at any point, and let its velocity be given in magnitude and direction by V. Then ρV is the current density of the electricity at the point in question. This expression may be made to include not only cases where the motion of a single charge or a small number of charges is contemplated, but also those like the flow of electricity along a wire in which the number of carriers is enormous. In such cases we have to take the average value of ρV, and it is quite clear that this is equal to the magnitude of the true electric current as specified on p. 97.

---

* *Ann. der Phys.* vol. CLVIII. p. 87 (1876).

When it is necessary or advisable to distinguish between the different kinds of electric current which may occur, we may write equations (4) in the form

$$\text{rot } H = 4\pi A \left(\rho V + \frac{dD}{dt}\right) \qquad \text{..................(7)}.$$

Where we are dealing with material systems we shall require the average, not the actual, values of these vectors just as in the theory of the behaviour of dielectric media.

Induction of Currents.

In an electrostatic field the work done in taking a unit charge round any closed path, to the point from which it started, vanishes. The electrostatic potential is a single-valued function of the space coordinates, and the electromotive force round any closed circuit is zero. This is no longer true in a region in which the magnetic induction is changing. Faraday* showed that when the magnetic field inside a closed conducting circuit was made to change, a current was caused to flow round the circuit. He also showed that the electromotive force round the circuit was equal to the rate of diminution of the flux of induction through the circuit in the positive direction, multiplied by a universal constant. This statement is very similar in form to that in which we expressed the First Law. If we put it into analytical form we shall see that the resemblance is very close indeed.

The electromotive force round any closed circuit is equal to the work done in taking a unit positive charge round the circuit. If s denotes length measured along the circuit this is

$$\int E \cos \hat{Es}\, ds = \int (E_x dx + E_y dy + E_z dz),$$

---

* *Exp. Res.* § 116.

where $E (= E_x, E_y, E_z)$ is the electric intensity. Let B be the magnetic induction at any point of any surface bounded by the circuit. The flux of induction through an element of area dS of this surface will be $B \cos \widehat{nB}\, dS$, where $\widehat{nB}$ is the angle between the resultant induction B and the normal to the surface at dS. But

$$B \cos \widehat{nB} = B_x \cos nx + B_y \cos ny + B_z \cos nz,$$

where $\cos nx$, $\cos ny$ and $\cos nz$ are the direction cosines of the normal. The total flux of induction through the surface is therefore

$$\iint (B_x \cos nx + B_y \cos ny + B_z \cos nz)\, dS,$$

and the rate of diminution of this is proportional to the electromotive force round the circuit, so that

$$\begin{aligned}&\int (E_x dx + E_y dy + E_z dz)\\ &\qquad = - A_1 \frac{d}{dt} \iint (B_x \cos nx + B_y \cos ny + B_z \cos nz)\, dS.\end{aligned}$$

But by Stokes's Theorem

$$\begin{aligned}&\int (E_x dx + E_y dy + E_z dz)\\ &\qquad = \iint \left\{ \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \cos nx + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \cos ny \right.\\ &\qquad\qquad\qquad \left. + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \cos nz \right\} dS.\end{aligned}$$

Since these relations are true for any surface bounded by the contour, the surface integrals must be identically equal; so that

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = - A_1 \frac{\partial B_x}{\partial t},$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = - A_1 \frac{\partial B_y}{\partial t},$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = - A_1 \frac{\partial B_z}{\partial t},$$

or

$$\operatorname{rot} E = - A_1 \frac{\partial B}{\partial t} \quad \text{......................(8).}$$

The value of the constant A_1 is determined by the units in which E and B are measured (see p. 111). These three equations represent the second group of Maxwell's equations. When it is necessary to give them a name we shall refer to them as the expression of the Second Law of Electrodynamics. Comparing them with the First Law,

$$\text{rot } H = 4\pi A \left(\rho V + \frac{\partial D}{\partial t}\right),$$

we see that, since B is proportional to H and D to E, there is a kind of reciprocal relation between them. Also they are not quite the same in form on account of the negative sign in front of $\frac{\partial B}{\partial t}$.

Not only do these equations enable us to deduce the whole of the phenomena of electromagnetism, but, as we shall see, they are the basis of the science of optics as well.

The Dynamical Theory of Electromagnetism.

The Second Law of Electromagnetism may be looked upon from two different standpoints according to the attitude we take towards electrical science. If we regard electrodynamics as more fundamental than dynamics proper, then we must regard the Second Law as a fundamental law of nature empirically given. We may however take the standpoint that the aether, which we postulate as a medium in which all electrical actions occur, will in the last analysis prove to be a mechanical system subject to the basal laws of dynamics. Provided we make this assumption, even though we know nothing of the nature of the mechanism, we can show that the Second Law is a consequence of the First Law. The view that electrical actions are ultimately dynamical is one whose development in the hands of Maxwell led to notable advances in the science, and it is the view towards which, at any rate until quite recently, most authorities have leaned. Nevertheless it is equally logical to accept the Second Law as an ultimate fact and then, later on, to consider what we can make of the laws of dynamics from the standpoint thus adopted.

We shall now proceed to consider some of the consequences which follow from the assumption that every electrodynamical system is a dynamical system subject to the operation of the first law and of the fundamental laws of electrostatics and magnetostatics. The energy of the field is equal to

$$\tfrac{1}{2}\kappa(E_x^2 + E_y^2 + E_z^2) + \tfrac{1}{2}\mu(H_x^2 + H_y^2 + H_z^2) \quad \text{......(9)}$$

per unit volume. Since a change in the electric intensity is accompanied by a corresponding change in the magnetic force, it is natural to assume that one of these terms represents potential and the other kinetic energy. Moreover the presence of magnetic energy is always associated with charges in motion whereas electrostatic energy is present when all the charges are at rest. The obvious conclusion therefore is that magnetic energy should be identified with kinetic energy and electrostatic with potential energy. A more rigorous proof of the necessity for identifying magnetic with kinetic energy will be found in Jeans's *Electricity and Magnetism*, p. 483.

Assuming that the magnetic energy of the electrodynamic field represents the kinetic energy of a dynamical system, let us consider the behaviour of a system of n circuits 1, 2, 3, etc. carrying currents i_1, i_2, i_3, etc. In this case the magnetic energy can be written as a quadratic function of the n variables i_1, i_2, i_3, etc. For the number of tubes of magnetic induction which flow through any one, let us say the mth, circuit, is the sum of n terms each representing a contribution from one circuit. This follows since the magnetic force at any point due to a current depends only on the relative geometry of the point and the current, the nature of the intervening medium, and the magnitude of the current, to which it is proportional. Thus if N_m is the number of tubes of magnetic induction which traverse the mth circuit in the positive direction

$$N_m = L_{1m} i_1 + L_{2m} i_2 + L_{3m} i_3 + \dots + L_{mm} i_m + \dots + L_{nm} i_n \quad \dots(10).$$

The coefficients L_{1m}, etc. depend only on the geometry of the circuits and the nature of the material in which they lie. They are called coefficients of self-induction when the suffixes are like, and of mutual induction when they are unlike.

The magnetic energy of any system, including that of a system of currents, is equal to $\frac{1}{2}\iiint \mu H^2 d\tau$, taken throughout the volume of the system. Let us suppose that the space is mapped out by means of unit tubes of induction. These are closed tubes which completely fill the space and never intersect. If δS is the normal sectional area of a unit tube at any point, the element of volume $d\tau$ may be replaced by $\delta S\, ds$, where ds is an element of length of the tube and is normal to δS. But δS is the area over which the flux of induction is unity, so that

$$\mu H\, \delta S = 1.$$

The part of the volume integral which belongs to any one tube is

$$\tfrac{1}{2}\int \mu H^2 \delta S ds = \tfrac{1}{2}\int H ds.$$

Thus if we indicate by Σ a summation over all the unit tubes, the magnetic energy T of the system becomes

$$T = \tfrac{1}{2}\iiint \mu H^2 d\tau = \tfrac{1}{2}\Sigma\int H ds.$$

Now by the first law of electrodynamics $\int H ds = 4\pi A\Sigma' i$, where $\Sigma' i$ is the algebraic sum of the currents which are embraced by the tube along which the integration is taken. Thus in the case of a system of linear currents T is equal to the sum, over all the tubes, of $2\pi A$ times the algebraic sum of the currents with which they are linked. This is equal to the sum, over all the current circuits, of $2\pi A$ times the number of tubes which are linked with each circuit. Hence

$$\begin{aligned} T &= 2\pi A\,\Sigma i_m N_m \\ &= 2\pi A\,\{L_{11}i_1^2 + 2L_{12}i_1 i_2 + \ldots + L_{nn}i_n^2\} \quad \text{.........(11)}, \end{aligned}$$

where N_m is the algebraic sum of the number of tubes which thread the mth circuit in the positive direction.

At first sight this result appears to be inconsistent with the conclusions which we formerly reached as to the equivalence of currents and magnetic shells. We see from equation (11), Chap. V, that the energy of the shell which is equivalent to the mth circuit is $-\phi_m \dfrac{N_m}{\mu}$, where N_m is the number of tubes of induction, supposed invariable, which thread it, ϕ_m is its strength, and μ is the permeability of the medium. For a medium of unit permeability $\phi_m = 4\pi A i_m$; so that for a medium of permeability μ, $\phi_m = 4\pi A\,\mu i_m$. This result follows since the magnetic *force* which a given current produces is independent of the medium, while it is the *induction* due to a shell which is independent of the permeability of the medium. Thus for a single equivalent shell in a fixed field the energy is $-4\pi A\, i_m N_m$. To find the total energy of the system of equivalent shells we have to imagine them created in infinitesimal steps, so that each increment is proportional to the final magnitude of the corresponding circuit. The calculation is precisely similar to that followed in

Chap. II, p. 33, in obtaining the total energy of a system of electrostatic charges, except for the difference in the form of the expression for the potential energy of the system. This calculation introduces a factor $\frac{1}{2}$, so that the energy of the system of shells which produces the same magnetic field as the system of currents is

$$-2\pi A\,\Sigma i_m N_m = -2\pi A\,(L_{11}i_1^2 + 2L_{12}i_1i_2 + \ldots + L_{nn}i_n^2).$$

The difference between this expression and (11) arises from the energy which lies within the volume of the shell itself*. This quantity, which is equal to $-2T$ in the case of the equivalent shells, is zero in the case of the currents.

The method of Lagrange's equations† enables us to find out a great deal about the behaviour of a mechanical system even when we have no means of discovering the precise nature of the mechanism. We shall therefore apply that method to the problem under consideration. As a preliminary we have to express the kinetic energy T and the potential energy W as functions of the generalized coordinates x_1, x_2, etc. and the corresponding velocities $\dot{x}_1$, $\dot{x}_2$, etc. We have seen that if we identify T with the magnetic energy in the field we get

$$T = +\,2\pi A\,(L_{11}i_1^2 + L_{12}i_1i_2 + L_{13}i_1i_3 + \ldots),$$

and this will be a quadratic function of the velocity coordinates $\dot{x}_1$, $\dot{x}_2$, etc. if we make $i_1 = \dot{x}_1$, $i_2 = \dot{x}_2$, $i_3 = \dot{x}_3$, etc. Thus the generalized displacement x_s becomes $\int i_s dt$ and is equal to the quantity of electricity which has flowed round the circuit after some fixed instant. In the present case $W = 0$ since the system does not possess electrostatic energy, the capacity being regarded as negligible. We also notice that the L's are functions only of the geometrical arrangement of the circuits and of the nature of the intervening medium. They do not involve the generalized displacements x.

If X_s is the component of generalized external force corresponding to the generalized displacement x_s, we have

$$\frac{\partial}{\partial t}\left(\frac{\partial T}{\partial \dot{x}_s}\right) - \frac{\partial T}{\partial x_s} = X_s \ldots\ldots\ldots\ldots\ldots\ldots\ldots (12),$$

---

* Cf. Jeans, *Electricity and Magnetism*, p. 433.

† For an account of the part of generalized dynamics which is germane to the present discussion the reader may be referred to Jeans's *Electricity and Magnetism*, Chap. XVI.

the motion being unconstrained. An equation of this type holds good for each generalized coordinate x_s whether the system is a conservative one or not.

[It may perhaps be worth while pointing out for the benefit of those readers who are unfamiliar with this branch of dynamics that the X's are not actually forces nor the x's actually displacements. The condition that they have to satisfy is that the work W done during a small change in the state of the system due to a change, let us say δx_s, of the value of the generalized coordinate x_s is $= X_s \delta x_s$. Thus in the present case, since δx_s means a change in the quantity of electricity which has flowed round the circuit, X_s will be the effective electromotive force in the circuit.]

Since $T = +\, 2\pi A\, [L_{11}\dot{x}_1{}^2 + L_{12}\dot{x}_1\dot{x}_2 + L_{13}\dot{x}_1\dot{x}_3 + \ldots]$,

$$\frac{\partial}{\partial t}\left(\frac{\partial T}{\partial \dot{x}_s}\right) = +\, 4\pi A\, \frac{\partial}{\partial t}\,(L_{1s}\dot{x}_1 + L_{2s}\dot{x}_2 + L_{3s}\dot{x}_3 + \ldots + L_{ss}\dot{x}_s + \ldots)$$

$$= +\, 4\pi A\, \frac{\partial N_s}{\partial t} \qquad \ldots\ldots\ldots\ldots(13),$$

where N_s is the number of tubes of induction which thread the sth circuit. As T does not contain the x's, $\dfrac{\partial T}{\partial x_s} = 0$, so that

$$+\, 4\pi A\, \frac{\partial N_s}{\partial t} = X_s.$$

Suppose that in this circuit there is an intrinsic electromotive force E_s; the work done during a small change δx_s in the quantity of electricity which has flowed round the circuit will be

$$X_s \delta x_s = E_s \delta x_s - R_s i_s{}^2 dt,$$

where R_s is the resistance of the circuit and i_s is the current round it. The last term represents the reaction of the matter in the circuit on account of the Joule heating effect. Since $i_s dt = \delta x_s$,

$$X_s = E_s - R_s i_s.$$

Thus
$$E_s - 4\pi A\, \frac{\partial N_s}{\partial t} = R_s i_s \qquad \ldots\ldots\ldots\ldots(14).$$

If there is no intrinsic electromotive force, E_s, in the circuit, the current is all due to the electromagnetic induction. From Ohm's Law we see that the electromotive force which arises from the electromagnetic induction is $R_s i_s = -4\pi A \dfrac{\partial N_s}{\partial t}$. Thus reverting to our former notation

$$\int (E_x dx + E_y dy + E_z dz) = -4\pi A \frac{\partial}{\partial t} \iint (lB_x + mB_y + nB_z)\, dS.$$

It follows that the Second Law is a consequence of the First Law provided we admit that the magnetic energy in the field can be identified with the kinetic energy of a mechanical system; and provided also that $A_1 = 4\pi A$.

The following investigation which leads directly to the second of Maxwell's equations is instructive. We shall suppose that there are both electrostatic and magnetic forces in the field, so that if T is the kinetic energy, and W the potential energy, per unit volume of the system, we have

$$T = \tfrac{1}{2}\mu (H_x^2 + H_y^2 + H_z^2),$$

and

$$W = \tfrac{1}{2}\kappa (E_x^2 + E_y^2 + E_z^2).$$

We shall restrict our proof to cases in which there are no free charges. Our system has to satisfy the first law of electrodynamics, so that

$$\text{rot}\, H = 4\pi A \frac{\partial D}{\partial t} = 4\pi A \kappa \frac{\partial E}{\partial t},$$

and we also have

$$\text{div}\, D = 0.$$

Now the changes occurring in every dynamical system are subject to the principle of Least Action, which may be put in the form

$$\delta \int_{t_1}^{t_2} (T - W)\, dt = 0 \quad \text{..................(15).}$$

This means that in any natural motion of the system, from a given configuration at time t_1 to another given configuration at time t_2, the actual motion is such as to make the time integral of $T - W$ a minimum, and that any slight variation from the actual motion of the system, subject to the conditions being unchanged for the initial and final configurations ($t = t_1$ and $t = t_2$ respectively), must be such that the variation of the integral is zero. We shall see that this is sufficient to establish the Second Law.

Since T is the kinetic energy, the H's will be velocity coordinates. Let us put $H = \dot{\theta}$ so that $H_x = \dot{\theta}_x$, $H_y = \dot{\theta}_y$, $H_z = \dot{\theta}_z$, where the θ's are generalized coordinates. We may now write the

First Law in the form $\operatorname{rot}\frac{\partial\theta}{\partial t} = 4\pi A\frac{\partial D}{\partial t}$, and we shall extend this equation so as to make it true of any change of the dependent variables θ and D whether in respect to the time or not. Thus if θ varies by $\delta\theta$ we shall suppose the variation to be conditioned by a variation δD of D according to the formula $\operatorname{rot}\delta\theta = 4\pi A\,\delta D$. We have, if $d\tau$ is an element of volume,

$$\delta\int d\tau\int_{t_1}^{t_2} T\,dt = \int d\tau\int_{t_1}^{t_2}\mu\,(\dot\theta_x\delta\dot\theta_x + \dot\theta_y\delta\dot\theta_y + \dot\theta_z\delta\dot\theta_z)\,dt$$

$$= \int \Big|\,\mu\,(\dot\theta_x\delta\theta_x + \dot\theta_y\delta\theta_y + \dot\theta_z\delta\theta_z)\,\Big|_{t_1}^{t_2} \times d\tau$$

$$- \iint_{t_1}^{t_2}\{\mu\,(\ddot\theta_x\delta\theta_x + \ddot\theta_y\delta\theta_y + \ddot\theta_z\delta\theta_z)\}\,d\tau\,dt,$$

and $$\delta\iint_{t_1}^{t_2} W\,d\tau\,dt = \int d\tau\int_{t_1}^{t_2}\kappa\,(E_x\delta E_x + E_y\delta E_y + E_z\delta E_z)\,dt$$

$$= \frac{1}{4\pi A}\int d\tau\int_{t_1}^{t_2}\left[E_x\left(\frac{\partial}{\partial y}(\delta\theta_z) - \frac{\partial}{\partial z}(\delta\theta_y)\right) + E_y\left(\frac{\partial}{\partial z}(\delta\theta_x) - \frac{\partial}{\partial x}(\delta\theta_z)\right)\right.$$

$$\left. + E_z\left(\frac{\partial}{\partial x}(\delta\theta_y) - \frac{\partial}{\partial y}(\delta\theta_x)\right)\right] dt$$

$$= \frac{1}{4\pi A}\int_{t_1}^{t_2} dt\int dS\,\Big[\delta\theta_x\,(nE_y - mE_z) + \delta\theta_y\,(lE_z - nE_x)$$

$$+ \delta\theta_z\,(mE_x - lE_y)\Big]$$

$$- \frac{1}{4\pi A}\int_{t_1}^{t_2} dt\int d\tau\left[\delta\theta_x\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y}\right) + \delta\theta_y\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z}\right)\right.$$

$$\left. + \delta\theta_z\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}\right)\right].$$

Thus

$$\delta\int d\tau\int_{t_1}^{t_2}(T - W)\,dt = 0 = \int d\tau\,\Big|\,\mu\,(\dot\theta_x\delta\theta_x + \dot\theta_y\delta\theta_y + \dot\theta_z\delta\theta_z)\,\Big|_{t_1}^{t_2}$$

$$- \frac{1}{4\pi A}\int d\tau\int_{t_1}^{t_2}\left\{\delta\theta_x\left(4\pi A\,\mu\ddot\theta_x + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\right.$$

$$+ \delta\theta_y\left(4\pi A\,\mu\ddot\theta_y + \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)$$

$$\left. + \delta\theta_z\left(4\pi A\,\mu\ddot\theta_z + \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\right\} dt$$

$$- \frac{1}{4\pi A}\int dS\int_{t_1}^{t_2}\Big[\delta\theta_x\,(nE_y - mE_z) + \delta\theta_y\,(lE_z - nE_x)$$

$$+ \delta\theta_z\,(mE_x - lE_y)\Big]\,dt.$$

The present analysis is restricted to the case where the media are at rest and where there are no free charges. Moreover the actual and the varied system are to be identical at $t = t_1$ and $t = t_2$. This makes the first volume integral vanish, since at both time limits $\delta\theta_x = \delta\theta_y = \delta\theta_z = 0$. The surface integral will vanish if we suppose the surface to be a closed surface at an infinite distance from the region in which the actions are going on. For although dS is proportional to r^2 both E and θ are of the order $\frac{1}{r^2}$ at most. It follows that the second volume integral must vanish. And since the $\delta\theta$'s are perfectly arbitrary both for each element of volume and for each element of time (except for the limits of time) the coefficients of each of them must always be zero. We therefore find

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -4\pi A\mu\ddot{\theta}_x = -4\pi A\frac{\partial B_x}{\partial t},$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -4\pi A\mu\ddot{\theta}_y = -4\pi A\frac{\partial B_y}{\partial t},$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -4\pi A\mu\ddot{\theta}_z = -4\pi A\frac{\partial B_z}{\partial t},$$

or

$$\operatorname{rot} E = -4\pi A\frac{\partial B}{\partial t}.$$

Thus the Second Law follows if we assume the truth of the First Law, and make the further assumption that the electrodynamic field is a mechanical system.

The foregoing deduction is practically identical with one given by Larmor (*Aether and Matter*, Chapter VI), who has shown that the laws of electrodynamics can be built up by giving the aether a mechanical constitution. For the further development of this theory, including the natural extension to the case where the presence in the field of electric charges is contemplated, the reader may be referred to Larmor's *Aether and Matter*, Chaps. VI, VII, and X.

Electrical Units.

We have seen that the work done in taking a unit magnetic pole round a path situated in free space which encircles a current of strength i once, is given by

$$\int H\,ds = 4\pi Ai.$$

The value of the universal constant A will clearly depend on the magnitude of the units in which H and i are measured. It also depends on the units of length and time.

We have already defined our units of magnetic pole strength and electric charge. From either of these units, although not from both, the units of all other electric and magnetic quantities may be derived unequivocally provided we keep to the same units of mass, length and time. Thus, to determine our unit of magnetic intensity H we only need to remember that mH is the force in dynes on a pole of strength m, so that the unit of H is determined by the unit of m and of force. It is evident that the dimensions of the product mH are independent of the dimensions which, since they are unknown, may arbitrarily be assigned to m; the product must always have the dimensions of a mechanical force. The dimensions of many other combinations of electric and magnetic quantities are predetermined in the same way; for instance, Ee has the dimensions of force and μH^2 and κE^2 have the dimensions of energy per unit volume, and so on. The unit of current i clearly only involves the units of electric charge and time, and is therefore determined in our case since we have already fixed the units of electric charge and time.

When H and i are measured in this way, the value of the constant factor is $4\pi A = 1/c$, where $c = 3 \times 10^{10}$ cms. per sec. The quantity c, which has the dimensions of a velocity, and thus has a numerical value which depends only on the units of length and time, is one of the most important physical quantities. As we shall see, it is equal, among other things, to the velocity of light and other forms of electromagnetic radiation in empty space.

When the stipulation above as to the character of the derived units is understood we can write equation (7) in the form

$$\text{rot } H = \frac{1}{c}\left(\frac{\partial D}{\partial t} + \rho V\right).$$

It remains to consider the value of the constant A_1 which enters into the expression for the second law. This may be discovered by making use of the principle of the conservation of energy, and for this purpose the simplest possible example of the induction of currents will suffice as well as another. Consider a single circuit carrying a current i, the self-induction of the circuit being L. Then the magnetic or kinetic energy of this circuit is $2\pi A\,Li^2 = \frac{L}{2c}i^2$. If the current i changes, the amount of this energy will alter and it is a known physical fact that there will be in consequence a self-induced electromotive force acting round the circuit. In general the circuit may lose energy by radiation or in other ways, but under suitable conditions these losses are negligible. In such cases the rate of loss of magnetic energy will be equal to the work done against the resistance of the circuit less the work supplied by the battery of electromotive force E. Thus

$$\frac{1}{2c}\frac{\partial}{\partial t}(Li^2) = Ei - Ri^2,$$

or

$$\frac{L}{c}\frac{\partial i}{\partial t} = E - Ri.$$

The electromotive force due to induction is therefore

$$E' = -\frac{L}{c}\frac{\partial i}{\partial t} = -\frac{1}{c}\frac{\partial N}{\partial t},$$

where N is the number of tubes of induction, measured in our system of units, which thread the circuit in the positive direction. It follows by comparison with the equations on p. 101 that $A_1 = \frac{1}{c} = 4\pi A$, when the electrostatic quantities are measured in the modified electrostatic units and the magnetic quantities in the modified electromagnetic units.

The following example is also instructive, since it brings out quite clearly that it must be the magnetic induction and not the magnetic intensity whose rate of change determines the magnitude of the induced currents. Consider the behaviour of a bar of magnetizable material encircled by a solenoid having n turns of wire per unit length. The bar is of uniform cross section a, and of indefinite length l. A current i flows in the solenoid and maintains a magnetic field of intensity H in the bar. The relation between H and i is

$$H = 4\pi Ani = \frac{n}{c}i \qquad \text{.....................(16).}$$

The work done in establishing the magnetic field in the bar consists of two parts, (1) $la\int H\,dI$, which, if I is the intensity of magnetization, represents the actual work in magnetizing the bar, and (2) $l\alpha\int H dH$, which represents the energy of the magnetic field in the aether of the space occupied by the bar. In a small interval of time dt the increment in the magnetic energy in the bar is thus

$$l\alpha H\left(\frac{dI}{dt}+\frac{dH}{dt}\right)dt.$$

This must be equal to the work done against the back electromotive force of induction. This work is

$$A_1 nl\alpha\frac{dB}{dt}\,.\,i\,dt = l\alpha H\,.\,A_1 c\,.\,\frac{dB}{dt}\,dt.$$

So that $$\frac{d}{dt}(I+H)=A_1 c\frac{dB}{dt}.$$

Thus $B=H+I$ in accordance with Chap. VI and $A_1=\frac{1}{c}=4\pi A$.

It is well to understand clearly the difference between the system of units we are using and the two systems, the electrostatic system and the electromagnetic system respectively, which are most frequently used in books dealing with the theory of electricity. The unit of electric charge in the electrostatic system is $\sqrt{4\pi}$ times our unit of electric charge; but it is not this difference so much as the difference in the units in which the magnetic quantities are measured which it is desirable to emphasize at the moment. On the electrostatic system the unit of current is obtained from the unit of electric charge, and the magnetic quantities are then obtained by giving A the arbitrary value unity in the equation

$$\int H\,ds = 4\pi A i.$$

This fixes the unit of magnetic force and so determines the unit of magnetic charge. In our units the measure of i is $\sqrt{4\pi}$ times greater and A is $4\pi c$ times less than in the electrostatic system, so that our unit of magnetic force is $c\sqrt{4\pi}$ times greater than the electrostatic unit. Since mH has the same value on all systems which have the same unit of mechanical force our unit of pole strength is $c\sqrt{4\pi}$ times smaller than the electrostatic unit.

The electromagnetic system of units also makes $A=1$, but it sets out by defining the unit magnetic pole as that which repels an equal pole at unit distance with a force of one dyne. On this system the unit of pole strength is therefore $\sqrt{4\pi}$ times our unit and $1/c$ times the electrostatic unit. The electrostatic unit of pole strength is thus greater than the electromagnetic unit by the factor 3×10^{10} cms. per sec. It follows that the measure of H on the electromagnetic system is $1/c$ times its measure on the electrostatic system. Whence, by considering the equation $\int H ds = 4\pi i$ which is true on both these systems, it follows that the measure of e on the electrostatic system is c times greater than on the electromagnetic system. Thus the electromagnetic unit of electric charge is c times the electrostatic unit, and $c\sqrt{4\pi}$ times our unit of electric charge.

It follows that when the unit of electric charge is defined as that which repels an equal and similar charge at unit distance with a force equal to $1/4\pi$ dynes, and the quantities which can be derived from it without making use of the two laws of electromagnetism are measured in terms of units which are based on this unit of electric charge: and when in addition the unit of magnetic pole strength is defined as the strength of that pole which repels an equal and similar pole at unit distance with a force of $1/4\pi$ dynes, and when the quantities which can be derived from this without making use of the two laws of electromagnetism are measured in units which are based on this unit of magnetic pole strength: then the two laws of electromagnetic induction become: .

(1) $$\operatorname{rot} H = \frac{1}{c}\left(\rho V + \frac{\partial D}{\partial t}\right) \quad \text{..................(17)},$$

(2) $$\operatorname{rot} E = -\frac{1}{c}\frac{\partial B}{\partial t} \quad \text{..........................(18)}.$$

In the sequel we shall always use (17) and (18) rather than (7) and (8).

Magnetic Force due to an Element of Electric Current.

We have seen (Chap. v, p. 85) that the force due to a magnetic shell of strength ϕ at a distant point P can be represented as arising from each element of the boundary of the shell. Since a current of strength i placed in a medium whose permeability is μ causes the same distribution of magnetic intensity as a shell of strength $\phi = 4\pi A \mu i = \mu i / c$, it follows that each element ids of an electric current can be regarded as giving rise at every point in space to a magnetic intensity of amount

$$\frac{i}{4\pi c} \frac{\sin \widehat{rs}\, ds}{r^2} \quad \text{......................(19)},$$

where r is the radius from the point P to ds and $\widehat{rs}$ is the angle between r and ds. The direction of the intensity is normal to r and ds. This result is true whether the space is empty or filled with homogeneous matter, and is independent of the permeability μ.

Force on an Element of Current in a Magnetic Field.

We have seen that the force on a magnetic shell of strength ϕ when placed in a magnetic field where the intensity is H can be represented as the resultant of forces equal to $-\phi [H ds]$ acting on each element ds of the boundary of the shell. It follows that the force exerted on an element ids of electric current i by a field H is

$$-\left[H . \frac{\mu i ds}{c}\right] = -\frac{1}{c}[B . ids] \quad \text{..............(20)}.$$

Where the current is due to the motion of an electric charge of volume density ρ with velocity v, the current per unit area is ρv, and the force on unit volume due to the field H is

$$-\left[\mu H . \frac{\rho v}{c}\right] \quad \text{......................(21)}.$$

If the charge is carried by a particle whose volume is τ the charge e will be given by $e = \iiint_\tau \rho d\tau$, and the force acting on it is

$$-\left[B . \frac{ev}{c}\right] \quad \text{......................(22)}.$$

The components F_x, F_y, F_z of this force are

$$\left.\begin{aligned} F_x &= \frac{e}{c}(B_z v_y - B_y v_z) \\ F_y &= \frac{e}{c}(B_x v_z - B_z v_x) \\ F_z &= \frac{e}{c}(B_y v_x - B_x v_y) \end{aligned}\right\} \quad \text{.................(23)}.$$

CHAPTER VII

ELECTROMAGNETIC WAVES

The Equations of Propagation.

LET us for the moment confine our attention to the application of Maxwell's equations to regions in which there are neither electric nor magnetic charges. Strictly speaking, from the point of view of the electron theory, this should restrict us to the case of the free aether, since all matter is supposed to be made up of electrically charged particles. In material media, however, the mean density of the charge at any point, if we average over a volume containing an enormous number of particles, is in most cases zero; so that we shall examine the consequences, incidentally, of supposing that with material media the peculiarities of the individual electrons can be left out of consideration. We shall soon see that the results at which we arrive, while exact for the free aether, are the crudest kind of approximation when applied to material media.

The equations to which we have been led in the preceding chapter in the case in which $\rho = 0$ are

$$\text{rot } H = \frac{1}{c}\frac{\partial D}{\partial t} = \frac{\kappa}{c}\frac{\partial E}{\partial t},$$

$$\text{rot } E = -\frac{1}{c}\frac{\partial B}{\partial t} = -\frac{\mu}{c}\frac{\partial H}{\partial t}.$$

We have seen that the two foregoing equations are really an abbreviation for six equations between the six components of H and E. These equations can therefore be solved for each one of the dependent variables H_x, H_y, H_z, E_x, E_y and E_z. Writing them out in full, we have

$$\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}=\frac{\kappa}{c}\frac{\partial E_x}{\partial t} \quad \text{.........................(1)},$$

$$\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}=\frac{\kappa}{c}\frac{\partial E_y}{\partial t} \quad \text{.........................(2)},$$

$$\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=\frac{\kappa}{c}\frac{\partial E_z}{\partial t} \quad \text{.........................(3)},$$

$$\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}=-\frac{\mu}{c}\frac{\partial H_x}{\partial t} \quad \text{.....................(4)},$$

$$\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}=-\frac{\mu}{c}\frac{\partial H_y}{\partial t} \quad \text{.....................(5)},$$

$$\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}=-\frac{\mu}{c}\frac{\partial H_z}{\partial t} \quad \text{.....................(6)}.$$

From (1), (5) and (6) we have

$$\begin{aligned}\frac{\kappa}{c}\frac{\partial^2 E_x}{\partial t^2}&=\frac{\partial}{\partial y}\left(\frac{\partial H_z}{\partial t}\right)-\frac{\partial}{\partial z}\left(\frac{\partial H_y}{\partial t}\right)\\ &=\frac{c}{\mu}\left\{\frac{\partial}{\partial y}\left(\frac{\partial E_x}{\partial y}-\frac{\partial E_y}{\partial x}\right)-\frac{\partial}{\partial z}\left(\frac{\partial E_z}{\partial x}-\frac{\partial E_x}{\partial z}\right)\right\}\\ &=\frac{c}{\mu}\left\{\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)E_x-\frac{\partial}{\partial x}\left(\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}\right)\right\}.\end{aligned}$$

Now $$\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=\rho=0,$$

since there are no charges in the medium. Thus

$$\nabla^2 E_x=\frac{\kappa\mu}{c^2}\frac{\partial^2 E_x}{\partial t^2}.$$

In the same manner we may show that each of the variables E_x, E_y, E_z, H_x, H_y, H_z satisfies the same equation.

We shall now investigate the solution of the equation

$$\frac{\partial^2 u}{\partial t^2}=a^2\nabla^2 u \quad \text{..........................(7)},$$

and show that it is the general equation of propagation of waves with velocity a. Let us take the integral of both sides of equation (7) throughout a closed volume, then

$$\iiint\frac{\partial^2 u}{\partial t^2}d\tau=a^2\iiint\nabla^2 u\,d\tau=-a^2\iint\frac{\partial u}{\partial n}dS \text{.........(8)}$$

by Green's Theorem (Chap. II, p. 24), where the surface integral extends over the surface enclosing the volume. Let S be the surface of a sphere of radius r whose centre is at the point P, and let $d\omega$ be the solid angle subtended by dS at the point P, then

$$-\iint \frac{\partial u}{\partial n}\, dS = \iint \frac{\partial u}{\partial r}\, r^2 d\omega = r^2 \frac{\partial}{\partial r} \iint u_r d\omega,$$

and
$$\iiint \frac{\partial^2 u}{\partial t^2}\, d\tau = \iiint \frac{\partial^2 u}{\partial t^2}\, r^2 d\omega\, dr = \frac{\partial^2}{\partial t^2} \iint d\omega \int_0^r r^2 u_r dr.$$

Differentiating both sides of (8) by the upper limit r,

$$r^2 \frac{\partial^2}{\partial t^2} \iint u_r d\omega = a^2 \frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r} \iint u_r d\omega\right).$$

Now $\iint u_r d\omega$ is 4π times the mean value of u over the surface of the sphere whose centre is P and radius r. Let us denote this by $\overline{u_r}$, then

$$r^2 \frac{\partial^2}{\partial t^2} \overline{u_r} = a^2 \frac{\partial}{\partial r}\left(r^2 \frac{\partial \overline{u_r}}{\partial r}\right),$$

or
$$\frac{\partial^2}{\partial t^2}(r\overline{u_r}) = a^2 \frac{\partial^2}{\partial r^2}(r\overline{u_r}).$$

Now let $r\overline{u_r} = v$ and introduce new independent variables $p = at + r$ and $q = at - r$, then

$$\frac{\partial v}{\partial t} = \frac{\partial v}{\partial p}\frac{\partial p}{\partial t} + \frac{\partial v}{\partial q}\frac{\partial q}{\partial t} = a\left(\frac{\partial v}{\partial p} + \frac{\partial v}{\partial q}\right),$$

$$\frac{\partial^2 v}{\partial t^2} = a^2\left(\frac{\partial^2 v}{\partial p^2} + 2\,\frac{\partial^2 v}{\partial p \partial q} + \frac{\partial^2 v}{\partial q^2}\right),$$

$$\frac{\partial v}{\partial r} = \frac{\partial v}{\partial p}\frac{\partial p}{\partial r} + \frac{\partial v}{\partial q}\frac{\partial q}{\partial r} = \frac{\partial v}{\partial p} - \frac{\partial v}{\partial q},$$

$$\frac{\partial^2 v}{\partial r^2} = \frac{\partial^2 v}{\partial p^2} - 2\,\frac{\partial^2 v}{\partial p \partial q} + \frac{\partial^2 v}{\partial q^2},$$

so that
$$\frac{\partial^2 v}{\partial p \partial q} = 0.$$

The general solution of this equation is clearly

$$\begin{aligned} v &= f_1(p) + f_2(q) \\ &= f_1(at + r) + f_2(at - r) = r\overline{u_r}, \end{aligned}$$

where f_1 and f_2 are arbitrary functions.

Since $\overline{u_r}$ is never infinite, $v = 0$ when $r = 0$, so that

$$f_2(at) = -f_1(at) \text{ for all values of } t.$$

Therefore f_1 and f_2 are not entirely arbitrary, but one is opposite in sign to the other. Thus

$$r\overline{u_r} = f_1(at+r) - f_1(at-r) \quad \text{..............(9).}$$

Differentiating by r

$$\overline{u_r} + r\frac{\partial \overline{u_r}}{\partial r} = f_1'(at+r) + f_1'(at-r).$$

When $r=0$, $\qquad \overline{u_0} = 2f_1'(at).$

But $\overline{u_0}$ is the mean value of u over a sphere of infinitesimal radius about the point P; it is therefore equal to the value u_P at P itself. So that

$$u_P = 2f_1'(at).$$

Again we have from (9)

$$\frac{\partial}{\partial r}(r\overline{u_r}) = f_1'(at+r) + f_1'(at-r),$$

and $$\frac{1}{a}\frac{\partial}{\partial t}(r\overline{u_r}) = f_1'(at+r) - f_1'(at-r).$$

Thus $$\frac{\partial}{\partial r}(r\overline{u_r}) + \frac{1}{a}\frac{\partial}{\partial t}(r\overline{u_r}) = 2f_1'(at+r)$$

and for $t=0$

$$\begin{aligned} 2f_1'(r) &= \left[\frac{\partial}{\partial r}(r\overline{u_r}) + \frac{1}{a}\frac{\partial}{\partial t}(r\overline{u_r})\right]_{t=0} \\ &= \left[\frac{\partial}{\partial r}\left(\frac{r}{4\pi}\iint u_r\,d\omega\right) + \frac{1}{a}\frac{r}{4\pi}\iint\frac{\partial u_r}{\partial t}\,d\omega\right]_{t=0}. \end{aligned}$$

Suppose that for a certain instant, which we shall take as the origin of time, the values of u_r and $\frac{\partial u_r}{\partial t}$ are given for every point in space. Let

$$(u)_{t=0} = F(x, y, z) \quad \text{and} \quad \left(\frac{\partial u}{\partial t}\right)_{t=0} = F'(x, y, z).$$

Then $$2f_1'(r) = \frac{\partial}{\partial r}\left(\frac{r}{4\pi}\iint F_r\,d\omega\right) + \frac{r}{4\pi a}\iint F_r'\,d\omega.$$

But when $r=at$, $2f_1'(r) = u_P$, so that

$$u_P = \frac{\partial}{\partial(at)}\left(\frac{at}{4\pi}\iint F_{at}\,d\omega\right) + \frac{t}{4\pi}\iint F'_{at}\,d\omega \quad \text{......(10).}$$

Thus the value of u_P at any time t subsequent to the time $t=0$, at which the values of F and F' are given throughout space, is obtained by describing a sphere of radius at about the point P.

The value of u_P will then be determined, provided we know the average value of $\frac{\partial u}{\partial t}$ over the surface of this sphere at $t=0$, as well as the rate at which the average value of u over this sphere at time $t=0$ changes as the radius is increased.

The meaning of this solution is perhaps seen more clearly if we consider a case in which the value of u at $t=0$ is zero everywhere except in a certain limited region of space τ. Consider the value of u at some external point P. Let the shortest line from P to τ be b and the longest line c. Then until $t=b/a$, the value of u_P will be zero since the sphere of radius at will not intersect the region for which u was not equal to zero at $t=0$. For similar reasons when t is greater than c/a the value of u_P will again be zero. Thus u represents a disturbance which is propagated in all directions with the velocity a. The foregoing solution, which is due to Poisson, shows that electromagnetic effects are propagated with finite velocity, like waves. In fact the equation

$$\frac{\partial^2 u}{\partial t^2}=a^2\nabla^2 u$$

is the general equation of wave motion and contains the mathematics underlying all the different kinds of wave motion contemplated in physics.

Before leaving Poisson's solution it may be well to consider a very simple concrete example to which it may be applied. Suppose that by means of two current sheets AB and CD perpendicular to the plane of the paper, we produce a uniform magnetic field H in the direction of the arrow. Suppose that at a certain time $t=0$ the currents are stopped. What will be the value of the magnetic intensity at the point P afterwards? Clearly there will be no field at P until $t=\frac{PR}{a}=\frac{b}{a}$, where a is the velocity of propagation; that is to say, until the sphere of radius at cuts the plane CD. To find the value after the sphere has

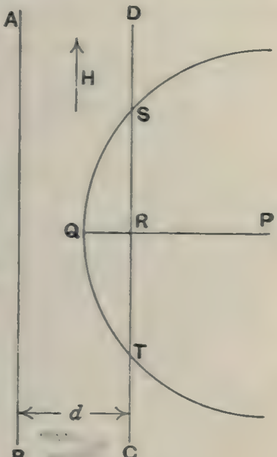


Fig. 22.

cut the plane CD but not the plane AB, let the points of intersection with CD be S and T. Then at $t=0$, $\frac{\partial H}{\partial t}=0$ everywhere except in a circular [illegible] ~~strip~~ of infinitesimal width of which ST is the diameter. Even in this infinitesimal strip $\frac{\partial H}{\partial t}$ is finite, so that the integral in (10) which involves F' is zero when $t=0$. Also $H = \text{constant} = F$ at every point of the spherical cap of which SQT is a section, and $\frac{1}{4\pi}\iint F\,d\omega = \frac{H \times QR}{2PQ} = H\,\frac{at-PR}{2at}$. Thus from equation (10) $H_P = \frac{H}{2}$, and is independent of the position of Q so long as it lies between the planes. When the sphere of radius at intersects both planes $\frac{1}{4\pi}\iint F\,d\omega = \frac{H}{2}\,\frac{d}{at}$, where d is the distance between the planes, and

$$H_P = \frac{\partial}{\partial (at)}\left(\frac{at}{4\pi}\iint F\,d\omega\right) = 0.$$

It is clear that the slab of thickness d in which the magnetic intensity is H splits into two slabs of equal thickness in which the magnetic intensity is $H/2$, and these are propagated in opposite directions normal to the faces of the slab with the velocity a.

The moving slabs of magnetic force form only part of the solution of the problem. This is clear because the energy of the magnetic field in the two moving slabs taken together is only half of the energy of the original magnetic field. It is necessary to consider the electric intensity E as well.

At $t=0$, $E=0$ everywhere, but $\frac{\partial E}{\partial t}$ is not zero. Consider any rectangle δx, δy in the plane of the paper such that δx crosses CD to which δy is parallel. Then, by the first law of electrodynamics, confining ourselves to the case of propagation in free space for which $\kappa=1$, $\mu=1$ and $a=c$,

$$\frac{1}{a}\iint \frac{\partial E}{\partial t}\,\delta x\,\delta y = \int H\,ds,$$

or

$$\frac{1}{a}\int \frac{\partial E}{\partial t}\,\delta x = H,$$

since everything is independent of y and the work $H\delta y$ vanishes along the right-hand side of the rectangle where $H=0$. But if the angle $SPQ=\theta$,

$$\frac{\partial E}{\partial t}\,d\omega=\frac{\partial E}{\partial t}\frac{2\pi at\sin\theta\,at\,d\theta}{a^2t^2}=-2\pi\frac{\partial E}{\partial t}\,d(\cos\theta)=\frac{2\pi}{at}\,\frac{\partial E}{\partial t}\,dx.$$

Thus $\iint\frac{\partial E}{\partial t}\,d\omega=\frac{2\pi}{at}\int\frac{\partial E}{\partial t}\,\delta x$ and, from (10),

$$E_P=0+\frac{t}{4\pi}\iint\frac{\partial E}{\partial t}\,d\omega=\frac{1}{2a}\int\frac{\partial E}{\partial t}\,dx=\frac{H}{2}.$$

When the sphere of radius PQ cuts both planes we see that the value of $\int\frac{\partial E}{\partial t}\,\delta x$ contributed by the second plane is equal and opposite to that from the first, so that E_P again becomes zero. Thus the slab of moving magnetic force H is accompanied by an equal electric force at right angles to and of equal magnitude with H. The electric force in the slab which moves to the right is in the opposite direction to that in the slab which moves to the left, whereas the magnetic force is in the same direction in both cases.

Velocity of Propagation.

We have seen that the components of the electric and magnetic intensities satisfy the equation $\frac{\partial^2 u}{\partial t^2}=a^2\nabla^2 u$, so that their changes in time and space are such as would arise if they were propagated from every point with velocity $V=c/\sqrt{\mu\kappa}$. For a vacuous space $\mu=1$ and $\kappa=1$, so that it follows that electromagnetic disturbances are propagated in vacuo with a velocity which is equal to c, the ratio of the electromagnetic unit of charge to the electrostatic unit. At the time that this conclusion was first reached (by Maxwell) it had not been shown that electromagnetic disturbances were propagated with finite velocity, so that there was no experimental material available, by means of which the conclusion could be quantitatively tested. Maxwell however put forward the view, which Faraday's instinct had previously led him to express although his mathematical limitations had probably prevented him from being able to deduce adequate experimental evidence in favour of it, that light itself was really an electromagnetic phenomenon. In that event, Maxwell predicted, its velocity of propagation should be equal to the ratio c of the two kinds of units. Maxwell himself carried out a determination of the value of c and found it to be equal to the velocity of light through space, within the limits of his experimental error. At a much later date (1888) Maxwell's Theory was put in a much stronger position by the researches of Hertz* who showed that vibrating electrical systems such as could be set up in the laboratory emitted electromagnetic waves. Hertz investigated the principal properties of these waves and showed that they were analogous to waves of light, from which they differed principally in the possession of much longer wave-lengths.

More recent determinations have shown a continually increasing agreement between the values of (1) the ratio of the two units of electric charge, (2) the velocity of electric waves and (3) the velocity of light in space. The agreement of the results of different observers is well exhibited by the following numbers which are taken from Jeans's *Electricity and Magnetism*, p. 506.

For the value of c, the ratio of the two units, the following results have been collected by H. Abraham† as likely to be most accurate:

Himstedt	$3\cdot0057 \times 10^{10}$	Abraham	$2\cdot9913 \times 10^{10}$
Rosa	$3\cdot0000 \times 10^{10}$	Pellat	$3\cdot0092 \times 10^{10}$
J. J. Thomson	$2\cdot9960 \times 10^{10}$	Hurmuzescu	$3\cdot0010 \times 10^{10}$
Perot and Fabry	$2\cdot9973 \times 10^{10}$		

The mean of these quantities is

$$c = 3\cdot0001 \times 10^{10} \text{ cms./sec.}$$

For the velocity of propagation of electromagnetic waves in air the following values are collected by Blondlot and Gutton‡:

Blondlot	$3\cdot022 \times 10^{10}$,	$2\cdot964 \times 10^{10}$,	$2\cdot980 \times 10^{10}$
Trowbridge and Duane			$3\cdot003 \times 10^{10}$
MacLean			$2\cdot9911 \times 10^{10}$
Saunders		$2\cdot982 \times 10^{10}$,	$2\cdot997 \times 10^{10}$

The mean of these quantities is

$$2\cdot991 \times 10^{10} \text{ cms./sec.}$$

---

* *Ann. der Phys.* vol. XXXIV. p. 551 (1888).

† *Rapports du Congrès de Physique*, Paris, 1900, vol. II. p. 267.

‡ *Rapports du Congrès de Physique*, Paris, 1900, vol. II. p. 283.

For the velocity of light in free aether Cornu* gives as the most probable value $3{\cdot}0013 \pm {\cdot}0027 \times 10^{10}$ cms./sec. Dividing by 1·000294 the refractive index of air referred to a vacuum, this gives for the velocity of light in air

$$3{\cdot}0004 \pm {\cdot}0027 \times 10^{10} \text{ cms./sec.}$$

The velocity of electric waves is known with much less accuracy than the other two quantities but they are undoubtedly all three identical in value within the limits of experimental error involved in each case.

Since the velocity of propagation is $c/\sqrt{\mu\kappa}$, the velocity of light in magnetic and dielectric media should be inversely proportional to the square root of the product of the magnetic permeability and the dielectric constant. Since it follows from the wave theory of light that the refractive index n of a medium is inversely as the velocity of propagation of the light through it, it follows that for different media of the same magnetic permeability

$$n^2 \propto \kappa.$$

This law has not been found to be even approximately verified for the waves which constitute light. In fact, a moment's consideration shows that it must be wrong, since it would make n constant, whereas the phenomenon of dispersion shows that n is a function of the wave-length. The fact that n^2 is not proportional to κ is not to be regarded as an objection to the electromagnetic theory of light. The theory on which it has been deduced is exact when applied to the free aether but its scope is not wide enough properly to account for the optical behaviour of material media. The reason for this is that material media contain electrically charged particles which are set into motion by the electric and magnetic forces of the light waves, and it is necessary to consider the dynamics of these particles to account satisfactorily for the optical behaviour of such media.

If we turn from light waves to the electrical vibrations of much lower frequency emitted by the Hertzian oscillator the state of affairs is very different. The period of these vibrations is, as a rule, great compared with the natural periods of the electrons in the molecules of the substance, so that the motion of the electrons is much the same as it would be under a steady field of the same magnitude as the instantaneous value of that due to the light wave. Under these circumstances the material can be treated as a continuous medium of definite dielectric coefficient κ, and for waves of this character the velocity of propagation in different media should be inversely as the square roots of the dielectric coefficients. This conclusion is substantiated by the results of experiments. Thus A. D. Cole* found the refractive index of water to be 8·9 whereas its dielectric coefficient $\kappa = 80$. With other substances the agreement appears to be satisfactory within the rather considerable limits of error of the determinations of the dielectric coefficients†.

---

* *Loc. cit.* p. 246.

We shall defer to the next chapter the consideration of the causes which make the behaviour of bodies towards light different from that predicted by the simple form of the electromagnetic theory which we have been discussing. There are, however, a number of phenomena exhibited by electromagnetic waves in their relation to matter which are partly true for light waves and strictly true for very long waves. The rest of this chapter will be occupied with an account of some of these.

Properties of a Plane-Polarized Electromagnetic Wave.

A solution of the equation $\frac{\partial^2 u}{\partial t^2} = a^2 \nabla^2 u$ is

$$u = u_0 e^{i\frac{2\pi}{\lambda}(at - lx - my - nz)} \quad \text{..................(11)}$$

provided $l^2 + m^2 + n^2 = 1$. The expression on the right-hand side is a complex quantity, being equal to

$$u_0\left[\cos\frac{2\pi}{\lambda}(at - lx - my - nz) + i\sin\frac{2\pi}{\lambda}(at - lx - my - nz)\right].$$

The real part of u therefore represents a disturbance of wave-length λ and amplitude u_0 which is propagated along the straight line $x/l = y/m = z/n$ with constant amplitude u_0 and constant velocity a. It is thus the appropriate specification of a monochromatic train of plane waves of wave-length λ. If we take the direction of propagation to be along the axis of z we shall have $n = 1$ and u will be equal to $u_0 e^{i\frac{2\pi}{\lambda}(at-z)}$ and its real part to

---

* *Wied. Ann.* vol. LVII. p. 290 (1896).

† Fleming, *Principles of Electric Wave Telegraphy*, p. 320.

$$u_0 \cos \frac{2\pi}{\lambda}(at - z) \quad \text{.....................(12).}$$

Let the real part of u represent E_x, the x component of the electric intensity in the wave front, and consider the train of waves for which $E_y = E_z = 0$. The electric intensity in this train will be completely specified by the equations

$$E_x = \text{real part of } E_0 e^{i\frac{2\pi}{\lambda}(at-z)}$$

or

$$E_x = E_0 \cos \frac{2\pi}{\lambda}(at - z) \quad \text{..................(13).}$$

It is clear that any equation between functions of complex variables such as, for example,

$$F_1(u, iv) = F_2(x, iy)$$

involves the separate truth of the two equations

$$\text{Real part of } F_1 = \text{Real part of } F_2$$

and

$$\text{Imaginary part of } F_1 = \text{Imaginary part of } F_2,$$

otherwise $i = \sqrt{-1}$ would be equal to a real quantity, which is absurd. This principle effects considerable simplification in the working out of problems arising in connection with the propagation of waves, as it enables us to work with the complex solution and then pick out the real parts at the end of our calculations. The advantage of this lies in the fact that the complex equations are usually simpler than their real equivalents.

Suppose that we are dealing with the train of plane waves propagated along the axis of z. Each of the vectors E_x, E_y, E_z, H_x, H_y, H_z which serve to specify the state of the medium at any point at any instant must be of the form $u_0 e^{i\frac{2\pi}{\lambda}(at-z)}$. The values however are not independent but have to satisfy the six equations on p. 116, viz.

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \frac{\kappa}{c}\frac{\partial E_x}{\partial t},$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \frac{\kappa}{c}\frac{\partial E_y}{\partial t},$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \frac{\kappa}{c}\frac{\partial E_z}{\partial t},$$

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\mu}{c}\frac{\partial H_x}{\partial t},$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\mu}{c}\frac{\partial H_y}{\partial t},$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\mu}{c}\frac{\partial H_z}{\partial t}.$$

As each of the dependent variables is of the form

$$\text{const.} \times e^{i\frac{2\pi}{\lambda}(at-z)}$$

we have $$\frac{\partial}{\partial t} = i\frac{2\pi}{\lambda}a, \quad \frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0, \quad \frac{\partial}{\partial z} = -i\frac{2\pi}{\lambda} \quad \text{......(14)},$$

so that $$\left.\begin{array}{ll} \frac{\kappa a}{c}E_x = H_y, & -\frac{\mu a}{c}H_x = E_y \\ -\frac{\kappa a}{c}E_y = H_x, & \frac{\mu a}{c}H_y = E_x \\ E_z = 0, & H_z = 0 \end{array}\right\} \quad \text{............(15).}$$

The equations above may all have a constant of integration added to them, but this would mean merely the superposition of a state of things independent of the time. This can have no relation to wave motion, so that we may leave the constants of integration out of account. We observe that both E_z and H_z are equal to zero, so that there is no component either of electric or magnetic intensity in the direction of propagation of the wave. *The resultant electric intensity and also the resultant magnetic intensity lie in the wave front.* Also $E_x H_x + E_y H_y = 0$ so that the resultant electric intensity is at right angles to the resultant magnetic intensity. In addition $\mu(H_x^2 + H_y^2) = \kappa(E_x^2 + E_y^2)$ so that the electric energy in the wave is equal to the magnetic energy. In the free aether μ and κ are each equal to unity in our units so that the electric and magnetic intensities are equal.

If we choose the axis of x so that the electric intensity lies along it, then the vectors which specify the wave are

$$E_x = E, \quad E_y = 0, \quad E_z = 0, \quad H_x = 0, \quad H_y = \frac{c}{\mu a}E = \frac{\kappa a}{c}E, \quad H_z = 0.$$

The electric intensity will always remain along the axis of x and the magnetic intensity along the axis of y. The wave is thus plane polarized; since its properties are not the same in reference to two planes at right angles to each other passing through the direction of propagation. We shall see that the plane which is usually called the plane of polarization in optics is that which contains the direction of the magnetic vector and is perpendicular to the direction of the electric vector. (See p. 132.)

It is evident that as the wave moves past any fixed point O, the electric and magnetic intensities always remain proportional to one another and both vanish and also both have their maximum values simultaneously.

In the case of waves of sound the kinetic energy is measured by the squares of the time rate of change of the displacement of the particles constituting the medium, and the potential energy by the squares of the strains which depend upon the rate of change of the same displacements in space. Both the velocities and the strains travel together or are in phase just as the magnetic energy and the electric energy in electromagnetic waves are in phase. This may be regarded as another reason for identifying magnetic and electric energy with kinetic and potential energy respectively.

The following elementary method of deducing the velocity of an electromagnetic wave is instructive. Let the wave be propagated along the line Oz and let E_x be the x component of the electric intensity in the wave front. Consider the state of things when the wave front lies in the plane perpendicular to Oz which passes through the point O. Then the lines OQ and OU (Fig. 23) are in the wave front. Describe the rectangles $TOPRUS$ and $TOPVQW$, TP being small compared with PR and PV. TP, SR and WV are perpendicular to the wave front, PR and ST are parallel to OU, the axis of x, and PV and TW are parallel to OQ, the axis of y. Consider the work done in taking a unit magnetic pole round the rectangle $PTWV$. The work along the part $QVPO$ of the path vanishes; for the electrical disturbance which constitutes the wave has not yet reached this part of the path. The work along OT is equal and opposite to the work along WQ by symmetry. The net amount of work is equal to $TW \times H_y$. But this is equal to $1/c$ times the total current embraced by the path. In the present case the current is all of the displacement variety, so that it must be equal to the rate of increase of the flux of induction through the circuit, in the units we have been using. The change in the amount of induction through the circuit arises entirely from the forward motion of the wave so that, if a is the velocity of propagation, it will be equal to $\kappa E_x \times a \times TW$ per unit time. Thus

$$TW \times H_y = TW \times \frac{\kappa a}{c} E_x,$$

or

$$H_y = \frac{\kappa a}{c} E_x.$$

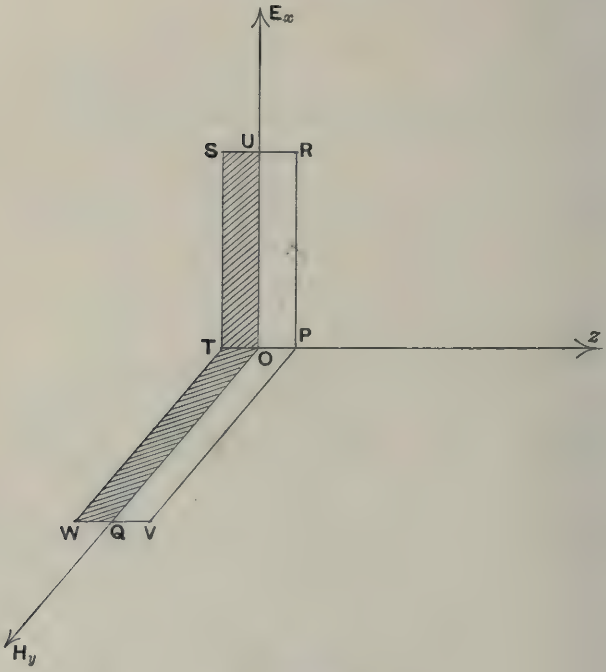


Fig. 23.

By considering the work done in taking a unit electric charge round the circuit $TPRS$ we find

$$E_x = \frac{\mu a}{c} H_y.$$

It follows that $a = c/\sqrt{\mu\kappa}$. We also see that if the electric force at any instant is along any particular direction, let us say the axis of x, it will always remain parallel to the axis of x, and the magnetic force will also always be parallel to the axis of y.

Reflexion and Refraction.

We shall now consider the behaviour of a train of plane waves incident on the boundary TPU between two insulating media whose dielectric constants and magnetic permeabilities are κ_1, μ_1 and κ_2, μ_2 respectively. The conditions which have to be satisfied by the components E_x, E_y, E_z of the electric intensity and the components H_x, H_y, H_z of the magnetic intensity at the surface of separation are:

(1) The tangential component of the electric intensity must be continuous in the two media.

(2) The normal component of the electric induction must be continuous in the two media.

(3) The tangential component of the magnetic force must be continuous in the two media.

(4) The normal component of the magnetic induction must be continuous in the two media.

Let us take the bounding surface to be perpendicular to the axis of z; then these conditions will be satisfied if, at every instant of time,

$$\left.\begin{array}{lll} E_{x_1}=E_{x_2}, & E_{y_1}=E_{y_2}, & \kappa_1 E_{z_1}=\kappa_2 E_{z_2} \\ H_{x_1}=H_{x_2}, & H_{y_1}=H_{y_2}, & \mu_1 H_{z_1}=\mu_2 H_{z_2} \end{array}\right\} \quad \text{.........(16)},$$

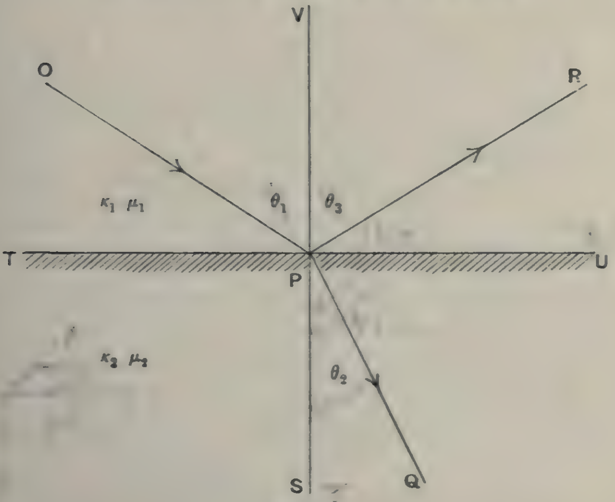


Fig. 24.

where the suffixes 1, 2 refer to the two media. These six equations are not independent. For if the first two are satisfied the sixth follows since the equation

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\mu}{c}\frac{\partial H_z}{\partial t}$$

must always hold in each medium. The third equation follows similarly from the fourth and the fifth by means of the relation

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \frac{\kappa}{c}\frac{\partial E_z}{\partial t}.$$

Thus only four of the six boundary equations of condition are really independent.

Wave polarized in the Plane of Incidence.

We shall now consider a wave propagated along OP and such that the resultant electric force E_x in the wave front is perpendicular to the plane of the paper. If the angle OPV is θ_1, PV being the normal to the boundary, the vectors which define this wave will all be proportional to

$$e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

where a_1 is the velocity and $\beta_1 = \frac{2\pi}{\lambda_1}$. They may therefore be represented by

$$E_x = X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)}, \quad E_y = 0, \quad E_z = 0;$$

$$H_x = 0, \quad H_y = \frac{c}{a_1\mu_1}\cos\theta_1 X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

$$H_z = -\frac{c}{a_1\mu_1}\sin\theta_1 X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

since these expressions satisfy equations (1)—(6).

The simplest way in which we can hope to satisfy the boundary conditions that the tangential electric intensity and the normal electric induction should be continuous is to make the y and z components of the electric intensity vanish for the refracted and reflected as well as the incident wave. The refracted wave will therefore be given by

$$E_x' = X_2 e^{\iota\beta_2(a_2 t - y\sin\theta_2 - z\cos\theta_2)}, \quad E_y' = 0, \quad E_z' = 0;$$

$$H_x' = 0, \quad H_y' = \frac{c}{a_2\mu_2} \cos\theta_2 X_2 e^{\iota\beta_2(a_2t - y\sin\theta_2 - z\cos\theta_2)},$$

$$H_z' = -\frac{c}{a_2\mu_2} \sin\theta_2 X_2 e^{\iota\beta_2(a_2t - y\sin\theta_2 - z\cos\theta_2)},$$

and the reflected wave by

$$E_x'' = X_3 e^{\iota\beta_3(a_3t - y\sin\theta_3 - z\cos\theta_3)}, \quad E_y'' = 0, \quad E_z'' = 0;$$

$$H_x'' = 0, \quad H_y'' = \frac{c}{a_3\mu_1} \cos\theta_3 X_3 e^{\iota\beta_3(a_3t - y\sin\theta_3 - z\cos\theta_3)},$$

$$H_z'' = -\frac{c}{a_3\mu_1} \sin\theta_3 X_3 e^{\iota\beta_3(a_3t - y\sin\theta_3 - z\cos\theta_3)}.$$

Since the boundary conditions have to be satisfied at all values of the time, the time factor in these expressions must be the same for all of them. Thus $\beta_1 a_1 = \beta_2 a_2 = \beta_3 a_3$. Also the velocity a is determined by the medium, so that $a_1 = a_3$. Hence $\beta_3 = \beta_1$ and $\beta_2 = \frac{a_1}{a_2}\beta_1$. The boundary conditions must be satisfied also for all values of y, so that the exponential factor in y must be common to each of the vectors. Thus $\beta_1 \sin\theta_1 = \beta_2 \sin\theta_2 = \beta_3 \sin\theta_3 = \beta_1 \sin\theta_3$. Therefore $\sin\theta_3 = \sin\theta_1$, so that the angle of reflexion must be equal in magnitude to the angle of incidence. Also

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\beta_2}{\beta_1} = \frac{a_1}{a_2} \quad \text{......................(17)}.$$

Thus the refractive index or the ratio of the sine of the angle of incidence to the sine of the angle of reflexion is equal to the ratio of the velocities of propagation of the light in the two media. Also with the convention as to the signs of a_1 and a_3 and of $\sin\theta_1$ and $\sin\theta_3$ which is here adopted, in conformity with general usage, $\cos\theta_3 = -\cos\theta_1$.

The boundary conditions will now be satisfied if

$$X_1 + X_3 = X_2 \quad \text{....................(18)},$$

$$\frac{c}{a_1\mu_1} \cos\theta_1 (X_1 - X_3) = \frac{c}{a_2\mu_2} \cos\theta_2 X_2 \quad \text{.........(19)},$$

$$\frac{c}{a_1} \sin\theta_1 (X_1 + X_3) = \frac{c}{a_2} \sin\theta_2 X_2 \quad \text{...........(20)}.$$

Since $\sin\theta_1/\sin\theta_2 = a_1/a_2$, equations (18) and (20) are identical. Solving equations (18) and (19) we find

$$X_2 = \frac{2\sin\theta_2\cos\theta_1}{\sin\theta_2\cos\theta_1 + \frac{\mu_1}{\mu_2}\sin\theta_1\cos\theta_2} X_1 \quad \text{.........(21)},$$

$$X_3 = \frac{\cos\theta_1\sin\theta_2 - \frac{\mu_1}{\mu_2}\sin\theta_1\cos\theta_2}{\cos\theta_1\sin\theta_2 + \frac{\mu_1}{\mu_2}\sin\theta_1\cos\theta_2} X_1 \quad \text{.........(22)}.$$

For all media on which experiments have been made which would serve to test these equations we may put $\mu_1 = \mu_2$, so that

$$X_3 = \frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)} X_1 \quad \text{..................(23)}.$$

Since the intensity of the waves is equal to the energy per unit volume multiplied by the velocity of propagation, and since the magnetic energy in a wave is always equal to the electric energy, the ratio of the intensity of the reflected to that of the incident wave will be

$$\left(\frac{X_3}{X_1}\right)^2 = \left[\frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}\right]^2,$$

where $\sin\theta_1/\sin\theta_2 = a_1/a_2$. The values given by these expressions for the intensities of the reflected and the refracted ray in terms of the intensity of the incident ray agree satisfactorily with those found experimentally when the light is polarized in the plane of incidence. This shows that the plane of polarization is the plane which contains the magnetic force and not that which contains the electric intensity.

Wave polarized in a Plane perpendicular to the Plane of Incidence.

In this case the magnetic force is to be perpendicular to the plane of incidence (the plane of yPz). We may therefore represent the incident wave by

$$H_x = H_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)}, \qquad H_y = 0, \qquad H_z = 0;$$

$$E_x = 0, \qquad E_y = -\frac{c}{\kappa_1 a_1}\cos\theta_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

$$E_z = +\frac{c}{\kappa_1 a_1}\sin\theta_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)}.$$

The expressions for the refracted and reflected waves are obtained by replacing the suffix 1 by the suffixes 2 and 3 respectively. As before, the exponential factors must be identical when $z = 0$ since the boundary conditions are to hold for all values both of t and y. It follows that

$$a_3 = a_1, \ \sin\theta_3 = \sin\theta_1 \ \text{and} \ a_1/a_2 = \beta_2/\beta_1 = \sin\theta_1/\sin\theta_2.$$

From the condition that the tangential magnetic force is continuous, we have

$$H_1 + H_3 = H_2.$$

The continuity of the tangential electric force gives

$$\frac{\cos\theta_1}{\kappa_1 a_1}(H_1 - H_3) = \frac{\cos\theta_2}{\kappa_2 a_2} H_2,$$

and the continuity of the normal electric induction

$$\frac{\sin\theta_1}{a_1}(H_1 + H_3) = \frac{\sin\theta_2}{a_2} H_2.$$

As before, the first and last of these equations are identical. For $\mu_1 = \mu_2 = 1$ we have $\kappa_1/\kappa_2 = a_2^2/a_1^2$ so that the second equation may be written

$$\sin\theta_1 \cos\theta_1 (H_1 - H_3) = \sin\theta_2 \cos\theta_2 \,.\, H_2.$$

Solving for H_2 and H_3 we find

$$H_2 = \frac{2\sin 2\theta_1}{\sin 2\theta_1 + \sin 2\theta_2} H_1 \quad \text{.................(24)}$$

and

$$-H_3 = \frac{1 - \dfrac{\sin\theta_1 \cos\theta_1}{\sin\theta_2 \cos\theta_2}}{1 + \dfrac{\sin\theta_1 \cos\theta_1}{\sin\theta_2 \cos\theta_2}} H_1$$

$$= \frac{1 + \tan^2\theta_1 - \tan\theta_1(\cot\theta_2 + \tan\theta_2)}{1 + \tan^2\theta_1 + \tan\theta_1(\cot\theta_2 + \tan\theta_2)} H_1$$

$$= \frac{\tan(\theta_2 - \theta_1)}{\tan(\theta_2 + \theta_1)} H_1 \quad \text{.....................(25).}$$

When $\theta_1 + \theta_2 = \pi/2$, $\tan(\theta_1 + \theta_2) = \infty$; so that for the particular value of $\theta_1 = \theta_p = \pi/2 - \theta_2$, the intensity of the reflected wave is zero. This angle is known as the polarizing angle. When θ_1 exceeds θ_p, H_3 becomes negative so that there is a sudden change of phase at the polarizing angle.

These formulae are found to be obeyed by light waves polarized perpendicularly to the plane of incidence for all angles except those in the immediate neighbourhood of the polarizing angle. It has been shown by Airy[*] and Jamin[†] that although the amount of reflected light becomes very small in the neighbourhood of the polarizing angle it never actually vanishes, and the change of phase which takes place in that neighbourhood is gradual and not sudden. Drude[‡] has pointed out that these differences between theory and experiment disappear if we suppose that the transition from the one medium to the other is gradual and not discontinuous. If there is a layer of very small but finite thickness in which the properties of the medium gradually change from those of the first to those of the second substance, the amount of reflected light never quite vanishes and the change of phase becomes a gradual one. The layer of transition may be small compared with the wave-length of light. It is probable from the existence of other physical phenomena, such as surface tension and the tenacious retention of surface films of gas by solid substances, that such transition layers do occur.

Conducting Media.

In those cases in which the medium affected by an electromagnetic disturbance possesses electrical conductivity, the displacement current will not be the only current which is set up. There will also be the true current of density j. Thus the Maxwellian equations suitable for this case are

$$(1) \quad \text{rot } H = \frac{1}{c}\left(j + \frac{\partial D}{\partial t}\right),$$

$$(2) \quad \text{rot } E = -\frac{1}{c}\,\frac{\partial B}{\partial t}.$$

Now if the medium is isotropic and σ is its specific conductivity, $j = \sigma E$, so that the first equation may be written

$$\text{rot } H = \frac{1}{c}\left(\sigma + \kappa\frac{\partial}{\partial t}\right) E \quad \text{..................(26).}$$

---

[*] *Camb. Phil. Trans.* vol. IV. p. 219 (1832).

[†] *Ann. Chem. Phys.* (3), vol. XXIX. p. 263 (1850); *ibid.* vol. XXXI. p. 165 (1850).

[‡] *Lehrbuch der Optik*, Leipzig, 1900, p. 266.

Let us suppose that waves of frequency p impinge on the boundary of such a medium. We shall see that the boundary conditions of the problem can be satisfied by means of a transmitted and a reflected wave both of which have the same frequency p. The time variation of all the vectors can therefore be appropriately represented by the common factor $e^{\iota pt}$.

Let us first consider the propagation of a wave of this frequency inside the conducting medium. The electric and magnetic forces have to satisfy the equations

$$\text{rot } H = \frac{1}{c}\left(\sigma + \kappa\frac{\partial}{\partial t}\right) E,$$

and

$$\text{rot } E = -\frac{\mu}{c}\frac{\partial H}{\partial t}.$$

Since $\frac{\partial}{\partial t} \equiv \iota p$ for the waves under consideration, these equations are equivalent to

$$\text{rot } H = \frac{1}{c}\left(\frac{\sigma}{\iota p} + \kappa\right)\frac{\partial}{\partial t} E = \frac{\kappa_1}{c}\frac{\partial E}{\partial t}.$$

and

$$\text{rot } E = -\frac{\mu}{c}\frac{\partial H}{\partial t}.$$

Thus the results which we have obtained for a non-conducting medium will still hold if we replace the dielectric constant of the medium κ by the complex quantity $\kappa_1 = \kappa + \sigma/\iota p$. The components of the electric and magnetic vectors still satisfy the equation

$$\frac{\partial^2 u}{\partial t^2} = a^2 \nabla^2 u,$$

but the constant a^2 is now complex and equal to $c^2/\mu(\kappa + \sigma/\iota p)$. Confining ourselves to the case in which the waves are propagated along the axis of z, since everything is proportional to $e^{\iota pt}$, the equation of propagation for this case reduces to

$$\frac{\mu}{c^2}(\iota p\sigma - p^2\kappa)\, u = \frac{\partial^2 u}{\partial z^2}.$$

The solution of this, appropriate to plane waves, is

$$u = e^{\pm(\alpha+\iota\beta)z} e^{ipt} \cdot u_0$$

where

$$(\alpha + \iota\beta)^2 = \frac{\mu}{c^2}(\iota p\sigma - p^2\kappa),$$

so that
$$\left.\begin{aligned} \alpha^2 - \beta^2 &= -\frac{\mu p^2 \kappa}{c^2} \\ 2\alpha\beta &= \mu p \sigma / c^2 \end{aligned}\right\} \quad\ldots\ldots\ldots\ldots\ldots\ldots(27).$$

Restoring the time factor $e^{\iota pt}$, the complete expression for u is $e^{\pm\alpha z}\, e^{\iota(pt \pm \beta z)}$. This will represent two waves, one propagated in the positive and the other in the negative direction along the axis of z. Confining ourselves to the one propagated in the positive direction we see that the vectors which determine it are given by $u = e^{-\alpha z}\, e^{\iota(pt - \beta z)}$. This shows that the wave is periodic in time $t = 2\pi/p$ and distance $z = 2\pi/\beta$. Thus if λ is the wave-length in the conductor $\beta = 2\pi/\lambda$. The occurrence of the real exponential term $e^{-\alpha z}$ shows that the amplitude diminishes as the wave progresses, the law of diminution, or "absorption," being that the amplitude falls off in equal ratios whilst the distance covered suffers equal increments. The quantity α is usually called the coefficient of absorption, but sometimes also the coefficient of extinction. $1/\alpha$ is the distance in which the amplitude falls off to $1/e$ of its initial value.

Solving (27) for α and β we have

$$\left.\begin{aligned} \alpha &= \frac{1}{c}\sqrt{\frac{\mu p}{2}}\,\sqrt{\{\sqrt{\sigma^2 + \kappa^2 p^2} - \kappa p\}} \\ \text{and} \qquad \beta &= \frac{\mu p \sigma}{2\alpha c^2} \end{aligned}\right\} \quad\ldots\ldots\ldots\ldots(28).$$

The formula for α shows that the coefficient of absorption is higher the higher the conductivity σ of the material. In other words, the higher the conductivity of the medium the more rapidly is it able to transform the energy of the electromagnetic waves traversing it into energy of other forms. If in (28) we put for σ the value of the electrical conductivity deduced from measurements with direct currents, the values of α which result are higher than those which would be obtained from measurements of the transparency of metals for light in the visible spectrum. Thus in the case of copper, and sodium light, we have approximately: $c = 3 \times 10^{10}$, $\mu = 1$, $p = 3 \times 10^{15}$, $\sigma = 4\pi c^2 \times 6 \times 10^{-4}$, $\kappa = 1$ and $\alpha = 3{\cdot}3 \times 10^6$ cms.$^{-1}$, whereas the value of α deduced from optical measurements* is $2{\cdot}8 \times 10^5$ cms.$^{-1}$. The particular value assigned to κ in calculating α from (28) does not exert any important influence on the result. The formula also does not suggest the observed difference in colour between the incident and transmitted light. These discrepancies arise from the fact that the current is carried by discrete electrons, with the consequence that both σ and κ are functions of p. The reason for this will be made clearer in the sequel; we shall now consider the phenomena which attend the reflexion of light at a conducting surface.

---

* Cf. Drude, *Lehrbuch der Optik*, 1st ed., p. 338.

Metallic Reflexion.

The problem of metallic reflexion is very similar to that furnished by the case of reflexion at the boundary between two insulating media. The same conditions as to continuity of the tangential electric and magnetic forces and of the normal electric and magnetic inductions have to be satisfied in both cases. The difference arises from the conducting power of the metallic medium, and we have seen that the type of theory proper to an insulating medium accounts for the propagation of waves in a conducting medium if we replace the dielectric constant by the complex quantity $\kappa_2' - \iota\sigma_2/p$, where σ_2 is the conductivity of the medium, κ_2' its dielectric coefficient and p the frequency of the waves. It is natural therefore to see if the boundary conditions cannot be fitted by the method previously adopted, the only change made being that the real quantity κ_2 in the former problem is replaced by the complex quantity $\kappa_2' - \iota\sigma_2/p$, where κ_2' is the real dielectric coefficient. We shall consider here only the case of waves polarized in the plane of incidence. A more complete discussion may be found in Drude's *Lehrbuch der Optik*, p. 334.

The incident wave being

$$E_x = X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)}, \qquad E_y = 0, \qquad E_z = 0;$$

$$H_x = 0, \qquad H_y = \frac{c}{a_1\mu_1}\cos\theta_1 X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

$$H_z = -\frac{c}{a_1\mu_1}\sin\theta_1 X_1 e^{\iota\beta_1(a_1 t - y\sin\theta_1 - z\cos\theta_1)},$$

the refracted and reflected waves will be given by similar expressions with the suffix 1 replaced by the suffixes 2 and 3 respectively. We have $a_1^2 = a_3^2 = c^2/\mu_1\kappa_1$ and $a_2^2 = c^2/\mu_2\kappa_2$. Thus since κ_2 is complex so also is a_2. Since

$$\beta_2 a_2 = \beta_1 a_1 \quad\text{and}\quad \beta_1 \sin\theta_1 = \beta_2 \sin\theta_2,$$

β_2 and $\sin\theta_2$ are also complex. The factors involving y and t being common, the two independent equations yielded by the boundary conditions become

$$X_1 + X_3 = X_2$$

and
$$\frac{\cos\theta_1}{a_1\mu_1}(X_1 - X_3) = \frac{\cos\theta_2}{a_2\mu_2} X_2.$$

Thus
$$X_2 = \frac{2X_1}{1 + \dfrac{a_1\mu_1\cos\theta_2}{a_2\mu_2\cos\theta_1}} \quad\text{......................(29)},$$

and
$$X_3 = \frac{1 - \dfrac{a_1\mu_1\cos\theta_2}{a_2\mu_2\cos\theta_1}}{1 + \dfrac{a_1\mu_1\cos\theta_2}{a_2\mu_2\cos\theta_1}} X_1 \quad\text{..................(30)}.$$

Now
$$\frac{a_1\mu_1\cos\theta_2}{a_2\mu_2\cos\theta_1} = \sqrt{\frac{\mu_1\kappa_2}{\mu_2\kappa_1}}\,\frac{\cos\theta_2}{\cos\theta_1}$$

is complex, and may be put in the form $\alpha + \iota\beta$, where α and β are real. Then

$$\frac{X_3}{X_1} = \frac{1 - (\alpha + \iota\beta)}{1 + \alpha + \iota\beta}$$
$$= \frac{1 - \alpha^2 - \beta^2 - 2\iota\beta}{(1+\alpha)^2 + \beta^2}.$$

The complex fraction may be put into the form

$$\rho e^{\iota\epsilon} = \rho(\cos\epsilon + \iota\sin\epsilon),$$

where
$$\rho\sin\epsilon = \frac{-2\beta}{(1+\alpha)^2 + \beta^2},$$

$$\rho\cos\epsilon = \frac{1 - \alpha^2 - \beta^2}{(1+\alpha)^2 + \beta^2};$$

so that
$$\rho^2 = \frac{(1-\alpha)^2 + \beta^2}{(1+\alpha)^2 + \beta^2} = 1 - \frac{4\alpha}{(1+\alpha)^2 + \beta^2} \quad\text{..........(31)},$$

and
$$\tan\epsilon = \frac{-2\beta}{1 - \alpha^2 - \beta^2} \quad\text{....................(32)}.$$

Thus E_x'' will be the real part of

$$\rho X_1 e^{\iota\beta_1\left(a_1 t - y\sin\theta_1 + z\cos\theta_1 + \frac{\epsilon}{\beta_1}\right)} \quad\text{...............(33)},$$

where ρX_1, β_1, θ_1, a_1, and ϵ are all real quantities. There will therefore be a difference of phase ϵ between the incident and reflected waves.

Experimental measurements of the ratio of the intensity of the reflected to the intensity of the incident radiation for infra-red radiation, in the case of normal incidence, have been made by Hagen and Rubens*. In this case $\theta_1 = \theta_2 = 0$, so that

$$\frac{a_1 \mu_1 \cos \theta_2}{a_2 \mu_2 \cos \theta_1} = \sqrt{\frac{\mu_1 \kappa_2}{\mu_2 \kappa_1}},$$

and
$$\frac{\mu_1}{\mu_2}\left(\frac{\kappa_2'}{\kappa_1} - \iota \frac{\sigma_2}{p\kappa_1}\right) = (\alpha + \iota\beta)^2.$$

Hence
$$\alpha^2 - \beta^2 = \frac{\mu_1}{\mu_2}\frac{\kappa_2'}{\kappa_1},$$

$$2\alpha\beta = -\frac{\mu_1}{\mu_2}\frac{\sigma_2}{p\kappa_1}.$$

Even in the visible part of the spectrum σ_2/p is a large quantity, although p is of the order 10^{15} and μ_1, μ_2, κ_1 and κ_2' are all comparable with unity. Thus $\alpha\beta$ is much larger than $\alpha^2 - \beta^2$. To a first approximation we may therefore put

$$\alpha = -\beta = \sqrt{\frac{\mu_1 \sigma_2}{2\mu_2 p\kappa_1}} \quad \text{………………(34)},$$

and, from (31),
$$\rho^2 = 1 - \frac{2}{\alpha}$$

$$= 1 - 2\sqrt{\frac{2\mu_2 p\kappa_1}{\mu_1 \sigma_2}} \quad \text{………………(35)}.$$

The results of Hagen and Rubens are expressed in terms of the reflecting power R of a metal. This is defined as the percentage of the incident intensity which is found in the reflected beam. Thus in terms of our notation

$$R = 100\rho^2,$$

and
$$100 - R = 200 \times \sqrt{\frac{2\mu_2 p\kappa_1}{\mu_1 \sigma_2}} \quad \text{…………(36)}.$$

The following table exhibits a series of values of $(100 - R) \times \sqrt{\sigma_2}$ found by Hagen and Rubens for a number of metals. We observe that for all the metals investigated, although there is a considerable range in the value of $100 - R$, the product $(100 - R)\sqrt{\sigma_2}$ is very nearly constant for each particular wave-length. The constancy is, on the whole, more marked the longer the wave-length λ.

---

* *Sitz. der Kon. Akad. der Wissensch.* Berlin, p. 269 and p. 210 (1903); *Ann. der Phys.* vol. XI. p. 873 (1903); *Phil. Mag.* [6] vol. VII. p. 157 (1904).

Metals	$\sqrt{\sigma_2}$	$\lambda = 4 \times 10^{-4}$ cm. $(100 - R)\sqrt{\sigma_2}$	$\lambda = 8 \times 10^{-4}$ cm. $(100 - R)\sqrt{\sigma_2}$	$(100 - R)$	$\lambda = 12 \times 10^{-4}$ c $(100 - R)\sqrt{\sigma}$
Silver	7·85	14·9	9·8	1·15	9·0
Copper	7·56	20·6	10·6	1·6	12·1
Gold	6·43	21·9	17·4	2·15	13·8
Platinum	3·04	25·8	14·0	3·5	10 6
Nickel	2·92	23·9	13·6	4·1	12·0
Steel	2·24	27·3	15·7	4·9	11·0
Bismuth	0·916	(22·7)	(16·9)	(17·8)	(16·3)
Patent Nickel *P*	1·95	15·4	14·6	5·7	11·1
Patent Nickel *M*	1·71	14·8	11·1	7·0	12·0
Constantan	1·43	16·7	10·6	6·0	8·6
Rosse's Alloy	1·44	16·6	13·0	7·1	10·2
Brandes and Schünemann's Alloy...	1·22	15·7	12·3	9·1	11·1

The values of σ_2 are those at 18° C., which was approximately the temperature at which the experiments were carried out.

The numbers in the preceding table were obtained by a direct comparison of the reflected and the incident radiation. This was furnished by a Nernst filament and the different wave-lengths were obtained by dispersion through a fluorite prism for the range from 10^{-4} cm. to 8×10^{-4} cm. and a sylvite prism for the range 8×10^{-4} cm. to 14×10^{-4} cm. For longer wave-lengths this method was unsatisfactory owing to the small intensity of the radiation thus obtained. An indirect method was therefore used instead. This consisted in comparing the radiation, emitted normally, from surfaces of the different substances with the radiation from a "perfectly black body" at the same temperature (see Chap. XV). The theory of the emission of radiation from hot bodies shows that the radiation emitted by any surface is to the radiation emitted by a similar surface of a perfectly black body at the same temperature as the radiation absorbed by the same body is to the radiation incident upon it. Since the reflected radiation is equal to the incident radiation less the radiation absorbed, this method enables the reflecting power of metals to be determined with accuracy.

The following table gives the values of the product $(100-R)\sigma^{\frac{1}{2}}$, obtained by this method for the residual rays of wave-length $25{\cdot}5 \times 10^{-4}$ cm. which remain after successive reflexions from surfaces of fluorite. Considering the smallness of $100-R$ the agreement of the experimental and computed values is remarkably good.

Metals	Emission-power $I=(100-R)$ for $\lambda = 25{\cdot}5 \times 10^{-4}$ cm. and 170°		Product $(100-R)\sqrt{\sigma_{170}}$ taken from column 2	
	Computed by Formula (36)	Observed		
Silver	1·15	1·13	7·07	Mean 7·33
Copper	1·27	1·17	6·67	
Gold	1·39	1·56	8·10	
Aluminium	1·60	1·97	8·91	
Zinc	2·27	2·27	7·24	
Cadmium	2·53	2·55	7·29	
Platinum	2·96	2·82	6·88	
Nickel	3·16	3·20	7·33	
Tin	3·23	3·27	7·32	
Steel	3·99	3·66	6·62	
Mercury	7·55	7·66	7·33	
Bismuth	10·09	(25·6)	(18·3)	
"Rotgiss"	2·73	2·70	7·16	Mean 7·41
Manganin	4·69	4·63	7·16	
Constantan	5·05	5·20	7·43	
Patent Nickel *P*	3·77	4·05	7·77	
Patent Nickel *M*	4·28	4·45	7·53	

There are two results of this investigation which are of special interest. In the visible part of the spectrum it is known that the reflecting power of metals does not agree with the predictions of Maxwell's Theory in its simple form, so that the present experiments determine the boundary of the region where other considerations have to be taken into account. The experimental results are in accordance with the simple theory when the wave-length of the radiation is equal to $25{\cdot}5 \times 10^{-4}$ cm.

The other point relates to the magnetic qualities of metals. The computed values in the tables have been obtained by putting $\mu_1 = \mu_2 = 1$ and $\kappa_1 = \kappa_2' = 1$ for all the metals. We see that the agreement in the case of the magnetic metals and alloys such as nickel, steel and so on is just as good as in the case of the non-magnetic metals. It follows that for oscillations of the frequency of those experimented with, the magnetic metals behave as though they were non-magnetic.

CHAPTER VIII

DISPERSION, ABSORPTION AND SELECTIVE REFLEXION

Medium containing Electrons.

In the last chapter we have seen that there are a number of consequences of Maxwell's Electromagnetic Theory which are satisfied by electromagnetic waves of very long period, but which are far from being borne out by the waves of much higher frequency which constitute the visible and ultra-violet regions of the spectrum. We shall now show that the reason for this is that we have neglected the part played by the inertia of the electrons. When that is taken into account we shall see that the results of the experiments, even on waves of very high frequency, are in very satisfactory agreement with the consequences of electromagnetic theory.

We shall suppose the matter under consideration to be made up of a large number of similar units which we shall call molecules. Each molecule contains a considerable number of electrons. In the absence of external electric force the electrons are to be regarded as establishing themselves in fixed positions of stable equilibrium, or in configurations of stable orbital motion. In the presence of an external electric field the electrons will be drawn away from the positions of undisturbed equilibrium in a manner similar to that discussed in Chap. IV. Owing to the inertia of the electrons their behaviour in a field which varies with the time is not so simple as that in the case of a steady field, which we have already considered.

Our first concern will be with the behaviour of an ideal substance which in all probability is somewhat simpler in its constitution than any real substance occurring in nature. We shall suppose that under the action of an electric intensity E the different electrons in the molecule denoted by the suffixes $1, 2, \dots n$, undergo displacements $x_1, x_2, \dots x_n$ from the equilibrium configuration. These displacements are supposed to be all in the same direction, which is that of the applied intensity E. We shall suppose that when a displacement x_1, for example, takes place, there is a force of restitution called into play which is proportional and parallel to x_1, and independent of $x_2, x_3, \dots x_n$. This assumption that the force of restitution, which we shall denote by x_1/λ_1, is independent of $x_2, x_3, \dots x_n$, can only be justified as a rough approximation, since it is clear that each of the displacements $x_2, \dots x_n$ is equivalent to a doublet of moment $e_2x_2, e_3x_3, \dots e_nx_n$, where the e's are the charges of the respective electrons, and it is evident that each of these doublets will give rise to a force on the electron e_1 proportional to its moment. The case in which this approximation is not made will be considered later (p. 169).

In addition to the forces of restitution each of the displaced electrons may, in general, be acted on by one or more forces of each of the following types:

(1) The external impressed force. This arises from the external electric intensity E, and it might be thought at first sight that its magnitude would be given by $Ee_1, Ee_2, \dots Ee_n$. The case here is, however, identical with that discussed at the end of Chapter IV, and the argument pursued there shows that the value of the actual force of external origin acting on the electrons is given by $(E + \alpha P)e_1 \dots (E + \alpha P)e_n$, where P is the polarization of the medium, and α is a constant, which, if the medium is sufficiently symmetrical, is equal to one-third.

(2) There will also be forces of frictional type tending to oppose the motion. These forces may be represented by a term proportional to the velocity, which in the case of the sth electron for example we may denote by $-\beta_s\dot{x}_s$ where β_s is a constant.

The precise mechanism of these forces is not yet properly understood. We shall see later that when an electrified particle is accelerated it emits radiation, and the emission of this radiation gives rise to a reaction $\frac{e^2}{6\pi c^3}\ddot{x}$ (see p. 266). For a simple harmonic vibration proportional to e^{ipt} this is equivalent to a force

$-\frac{e^2p^2}{6\pi c^3}\dot{x}$, and may therefore be represented as a frictional term provided $\beta_s = \frac{e^2p^2}{6\pi c^3}$. It is likely enough that forces of this character will always be present in the case of vibrating electrons. In most cases however they seem to be incapable of accounting for more than a minute part of the term $-\beta_s\dot{x}_s$, as the observed absorption is, in general, much greater than that to which the frictional force $-\frac{e^2p^2}{6\pi c^3}\dot{x}$ would give rise.

Lorentz* has suggested that the frictional term arises through the disturbance of the motion of the oscillating particles caused by the impacts of the molecules. At each collision the molecule is regarded as so profoundly shaken up that the regular forced oscillations are converted into irregular motions. The average result will clearly be of the nature of an absorption of energy. In fact Lorentz finds (*loc. cit.*) that if τ is the average interval between two collisions, the average result is the same as if there were a frictional force equal to $-\frac{2m}{\tau}\dot{x}$ acting on the electrons, where m is their mass. Unfortunately this cause does not seem to be large enough to account for the observed effects. It is found that the values of τ calculated from absorption phenomena are much smaller than those deduced from the kinetic theory of gases.

It seems to the writer that the following view of the mechanism of the absorption of light has much to recommend it. In the majority of cases it is probable that the resistance to the motion due to radiation, and to the effect of intermolecular collisions is small and comparatively unimportant. It is probable that in the neighbourhood of one of the principal periods of the substance the absorption of energy by the electrons continues until the equilibrium of the vibrating system becomes unstable. If the energy absorbed is sufficiently great, the electron will be emitted from the molecule and photo-electric phenomena will be exhibited. Nevertheless this need not necessarily be the case. All that is necessary is that there should be a temporary rearrangement among the electrons inside the molecule. In either event the energy of the resonating system will be converted into energy of other types, and that particular degree of freedom will possess comparatively little energy when the original system is reformed. The general results of this view are in qualitative agreement with those which follow from the assumption of a retarding force proportional to $\dot{x}$. They show the same gradual change in phase in passing through an absorption band, and indicate a maximum absorption and a refractive index equal to unity, when the period ? of the light is approximately coincident with the natural period of the substance.

---

* *Theory of Electrons*, p. 141.

In view of the facts that absorption does occur, and that its mode of occurrence is still doubtful, we shall for the present content ourselves with the assumption of a term, in the expression for the force, proportional to $-\dot{x}$. This term is to be regarded, not as the expression of a fundamental truth, but as a simple and convenient mathematical approximation whose consequences simulate the observed effects.

(3) In general we shall have to take account of the effect of the presence of a magnetic field. We have seen (Chap. VI, p. ~~68~~) 74 that a magnetic field of intensity H acting on a moving electron, charge e_s, gives rise to a force whose components are

$$\frac{e_s}{c}(H_z \dot{y}_s - H_y \dot{z}_s),$$

$$\frac{e_s}{c}(H_x \dot{z}_s - H_z \dot{x}_s),$$

$$\frac{e_s}{c}(H_y \dot{x}_s - H_x \dot{y}_s).$$

The total force acting on the electron will be compounded of all the forces mentioned, and, in the most general case, its equations of motion will be

$$m_s \frac{\partial^2 x_s}{\partial t^2} = e_s (E_x + \alpha_s P_x) - \frac{x_s}{\lambda_s} - \beta_s \frac{\partial x_s}{\partial t} + \frac{e_s}{c}\left(H_z \frac{\partial y_s}{\partial t} - H_y \frac{\partial z_s}{\partial t}\right),$$

$$m_s \frac{\partial^2 y_s}{\partial t^2} = e_s (E_y + \alpha_s P_y) - \frac{y_s}{\lambda_s} - \beta_s \frac{\partial y_s}{\partial t} + \frac{e_s}{c}\left(H_x \frac{\partial z_s}{\partial t} - H_z \frac{\partial x_s}{\partial t}\right),$$

$$m_s \frac{\partial^2 z_s}{\partial t^2} = e_s (E_z + \alpha_s P_z) - \frac{z_s}{\lambda_s} - \beta_s \frac{\partial z_s}{\partial t} + \frac{e_s}{c}\left(H_y \frac{\partial x_s}{\partial t} - H_x \frac{\partial y_s}{\partial t}\right) \quad \text{......(1).}$$

There will be three equations similar to these, but with different coefficients, for each of the n classes of electrons. The solution of these equations under the most general conditions leads to very complicated expressions, and does not lead to any results of physical importance which cannot be obtained more simply otherwise. We shall therefore consider the operation of the different forces separately, in so far as they can be separated physically, and thus find out the kind of effect which arises from each cause. In particular the effect of the last two terms, which depend on the presence of a magnetic field, will be deferred to a later chapter, which deals with magneto-optical and spectroscopic phenomena. We shall also confine ourselves to the case in which plane waves, polarized in the plane of yOz, are propagated along the axis of z, since all other cases may be made to depend on this, if necessary.

No Friction.

It seems desirable to treat this case separately, although of course the same results are obtained if we include friction and then in the final expressions put the coefficients β equal to zero. The chief advantage of thus lengthening the treatment is that effects due to resonance are then sharply marked off from those due to absorption and are brought out more simply than when the two effects are considered together. In the case under consideration the equation of motion of the sth type of electron is

$$m_s \frac{\partial^2 x_s}{\partial t^2} = e_s (E_x + \alpha_s P_x) - \frac{x_s}{\lambda_s} \quad \text{..............(2)},$$

or, dropping the subscripts for the present,

$$\left(\frac{m}{e} D^2 + \frac{1}{\lambda e}\right) x = E_x + \alpha P_x$$
$$= (E + \alpha P)\, \epsilon^{\iota p t}$$

if the frequency of the electric waves is p and $D \equiv \frac{\partial}{\partial t}$. E and P are constants which represent the maximum electric intensity and electric polarization in the wave. Our equation is a linear differential equation of the second order and the complementary function is

$$A_1 \epsilon^{\frac{\iota}{\sqrt{\lambda m}} t} + A_2 \epsilon^{-\frac{\iota}{\sqrt{\lambda m}} t} \quad \text{..................(3)}$$

where A_1 and A_2 are arbitrary constants and $p_0 = (\lambda m)^{-\frac{1}{2}}$ is the frequency of the natural vibrations of this electron. The particular integral is

$$\left(\frac{m}{e} D^2 + \frac{1}{\lambda e}\right)^{-1} (E + \alpha P) \epsilon^{\iota pt}$$

$$= \frac{e}{m} \frac{(E + \alpha P) \epsilon^{\iota pt}}{p_0^2 - p^2} \quad \text{....................(4).}$$

The complete solution is the sum of the expressions (3) and (4). The arbitrary constants A_1 and A_2 are to be determined, in any particular case, by the given initial conditions. If x and $\dot{x}$ are zero when the waves begin to fall on the electron the natural vibrations $A_1 \epsilon^{\iota p_0 t} + A_2 \epsilon^{\iota p_0 t}$ have the same initial amplitude as the forced vibrations of frequency p, and in any event the amplitudes of the vibrations of frequencies p_0 and p are initially comparable. There is however a very marked difference in the behaviour of the two vibrations as time progresses. The energy of the vibrations of frequency p_0 will gradually disappear owing to radiation and after a time their amplitude will become negligible. On the other hand any loss of energy from the vibrations of frequency p is continually made good by the action of the waves which have the same period and the same or opposite phase. These conclusions can be established quite strictly if a small coefficient β of frictional type is introduced into the equations. It is then found that the complementary function contains a factor ϵ^{-at} where a is a positive constant, whereas there is no such term in the particular integral. This shows that after the lapse of a sufficient interval the vibrations will be represented by this particular integral only. The student is recommended to work out in detail the case in which there is a small frictional term and the initial conditions are, let us say, $x = 0$ and $\dot{x} = 0$.

It follows that when sufficient time has elapsed for the system to have got into a steady condition, the value of x_s is given by

$$x_s = \frac{e_s}{m_s} \frac{(E + \alpha P) \epsilon^{\iota pt}}{p_s^2 - p^2}$$

$$= \frac{e_s}{m_s} \frac{E_x + \alpha P_x}{p_s^2 - p^2}.$$

Let ν_s represent the number of particles of type s per unit volume. Then the polarization

$$P_x = \sum_{s=1}^{s=n} \nu_s e_s x_s = (E_x + \alpha P_x) \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}.$$

Therefore
$$P_x = \frac{\sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}}{1 - \alpha \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}} E_x,$$

and
$$D_x = E_x + P_x = \left[1 + \frac{\sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}}{1 - \alpha \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}}\right] E_x.$$

From the symmetry of equations (1) it is clear that in general

$$\frac{D_x}{E_x} = \frac{D_y}{E_y} = \frac{D_z}{E_z} = \frac{D}{E} = \kappa.$$

Now the equations which determine the propagation of the waves are

$$\text{rot } H = \frac{1}{c} \frac{\partial D}{\partial t} = \frac{\kappa}{c} \frac{\partial E}{\partial t},$$

and
$$\text{rot } E = -\frac{1}{c} \frac{\partial B}{\partial t} = \frac{\mu}{c} \frac{\partial H}{\partial t}.$$

The velocity of propagation is $c/\sqrt{\mu\kappa}$ and the refractive index $m = \sqrt{\mu\kappa}$. Since we may put $\mu = 1$ for practically all dispersive media, we have

$$m^2 = 1 + \frac{\sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}}{1 - \alpha \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}},$$

or
$$\frac{m^2 - 1}{m^2 - \frac{\alpha - 1}{\alpha}} = \alpha \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)} \quad \text{........................(5).}$$

For substances which are sufficiently isotropic (see p. 72, Chap. IV) the value of α is one-third. Formula (5) then becomes

$$\frac{m^2 - 1}{m^2 + 2} = \frac{1}{3} \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)}.$$

As the density of the medium is made to change, by compressing it for example, the only quantities on the right-hand side which are altered are the ν's, each of which is proportional to the density. The right-hand side may therefore be written in the form $C\rho$, where ρ is the density and C is a quantity which is constant for any given substance and wave-length, but which varies from one substance and wave-length to another. The formula

$$\frac{m^2-1}{m^2+2} = C\rho \quad\text{..........................(6)}$$

was first deduced empirically by L. Lorenz* and was subsequently shown to be a consequence of the electromagnetic theory by H. A. Lorentz†. The most exacting test to which the formula has been put is that of calculating the refractive index of vapours from that of the corresponding liquids. The following figures, given by Lorentz, show the extent of the agreement which may be obtained in some cases:—

Material	Liquid		Vapour		
				Refractive Index	
	Density	Refractive Index	Density	Observed	Calculated from (6)
Water	·9991	1·3337	·000809	1·000250	1·000250
Carbon Bi-sulphide	1·2709	1·6320	·00341	1·00148	1·00144
Ethyl Ether	·7200	1·3558	·00332	1·00152	1·00151

It is to be borne in mind that comparisons of the kind here discussed are always to be made with light of the same wave-length.

It will be observed that in the deduction of formula (5) there is nothing which compels us to consider the medium as a chemically simple substance. Provided α has the same value for the electrons in all the different kinds of molecules present the value of

---

* *Ann. Phys. Chem.* vol. II. p. 70 (1880). A rather similar formula (see p. 155) had been derived from theoretical considerations much earlier by Maxwell, *Cambridge Calendar* (1869). Cf. Lord Rayleigh, *Phil. Mag.* vol. XLVIII. p. 151 (1899), and Sellmeier, *Ann. der Phys.* vol. CXLV. pp. 399, 520 (1872).

† *Arch. Néerl.* vol. XXV. p. 525 (1892).

$$\frac{m^2-1}{m^2-\frac{\alpha-1}{\alpha}}$$

will consist of the sum of a series of terms each of which consists of the product of two factors. The first factor is the number of particles of given type per unit volume and the second is a function of the frequency which has a definite value for each type of particle. Thus the value of

$$\frac{m^2-1}{\left(m^2-\frac{\alpha-1}{\alpha}\right)\rho}$$

for a mixture may be calculated by multiplying the value of this fraction which is characteristic of each of the substances present, by the proportion of the substance present in the mixture. Putting $\alpha=\frac{1}{3}$ this result may be written

$$\frac{m^2-1}{m^2+2}\frac{1}{\rho}=\frac{m_1^2-1}{(m_1^2+2)\rho_1}f_1+\frac{m_2^2-1}{(m_2^2+2)\rho_2}f_2+\frac{m_3^2-1}{(m_3^2+2)\rho_3}f_3+\ldots,$$

where m_s, ρ_s denote the values of m and ρ for the sth constituent present and f_s is the mass of it in unit mass of the mixture. This formula has been found to be fairly satisfactory for mixtures of different liquids.

It is found that this additive law of refractivity is not confined to merely physical mixtures. Something of the same kind holds for the individual atoms of which different bodies are composed. Supposing that α has the same value ($\frac{1}{3}$) for every electron no matter in what atom it may happen to be, expression (5) may be written

$$\frac{m^2-1}{m^2+2}=\frac{1}{3}\left[\sum_1^{n_a}\frac{a_s e_s^2}{m_s(p_s^2-p^2)}+\sum_1^{n_b}\frac{b_s e_s^2}{m_s(p_s^2-p^2)}+\ldots\right],$$

where a_s, b_s, etc. represent, for each atom A, B, etc., the number of electrons of given type per unit volume, and the separate summations are for each distinct kind of atom present. Thus $\sum_1^{n_a}\frac{a_s e_s^2}{m_s(p_s^2-p^2)}$ is to be taken over all the electrons in an atom of the first element A. If m_a is the refractive index of the element

A (supposed liquid) then, if the refractive power is independent of the state of chemical combination,

$$\frac{m_a^2-1}{m_a^2+2}=\frac{1}{3}\sum_1^{n_a}\frac{a_s' e_s^2}{m_s(p_s^2-p^2)},$$

where a_s' is now the number of electrons of type s present in unit volume of the liquid element. If M is the molecular weight of the compound, A, B, C ... are the atomic weights of the elements, and if a molecule of M contains q atoms of A, r of B, s of C and so on, then

$$M=qA+rB+sC+\dots.$$

If the density of the compound is ρ_m and the densities of the elements are ρ_a, ρ_b, ρ_c, etc., then

$$\frac{a_s}{a_s'}=q\frac{\rho_m}{M}\frac{A}{\rho_a},\qquad \frac{b_s}{b_s'}=r\frac{\rho_m}{M}\frac{B}{\rho_b},\text{ etc.}$$

So that

$$\frac{1}{3}\sum_1^{n_a}\frac{a_s e_s^2}{m_s(p_s^2-p^2)}=q\frac{\rho_m}{M}\frac{A}{\rho_a}\times\frac{1}{3}\sum_1^{n_a}\frac{a_s' e_s^2}{m_s(p_s^2-p^2)},\text{ etc.}$$

Thus

$$\frac{m^2-1}{m^2+2}\frac{M}{\rho_m}=q\frac{m_a^2-1}{m_a^2+2}\frac{A}{\rho_a}+r\frac{m_b^2-1}{m_b^2+2}\frac{B}{\rho_b}+\dots.$$

The quantity $\frac{m^2-1}{m^2+2}\frac{M}{\rho_m}$ is called the molecular refractivity of the compound and the quantities $\frac{m_a^2-1}{m_a^2+2}\cdot\frac{A}{\rho_a}$, $\frac{m_b^2-1}{m_b^2+2}\cdot\frac{B}{\rho_b}$, etc., are called the atomic refractivities of the elements.

The fact that the molecular refractivity of different substances can be calculated from the atomic refractivities is quite important. It shows that the mechanism which gives rise to refraction is an atomic property and is so deep-seated in the atom that it is almost uninfluenced by the changes of configuration of the atomic constituents which take place during chemical combination. This certainly seems to be fairly true in a large number of cases of chemical combination. It is not, however, universally true. Thus in the case of the compounds of carbon, the carbon atom is found to have a sufficiently constant atomic refractivity in all the so-called saturated compounds. If, however, two of the carbon atoms are united by a "double bond" it is necessary to assign to them a quite different atomic refractivity. Again, the value of the atomic refractivity of oxygen is quite different for hydroxyl oxygen and carbonyl oxygen. These several differences are so distinctive that the magnitude of the molecular refractivity is regarded as a valuable aid in determining the constitution of organic compounds. On the whole, the general constancy of the atomic refractivities, and the definiteness of the changes when changes do take place, supply us with striking evidence that the structure of the atoms is not very profoundly disturbed by chemical combination.

Dispersion.

We shall now consider the way in which, according to our formulae, the refractive index depends upon the frequency or wave-length of the incident vibrations. Formula (5) may be written somewhat more simply, in the form

$$\frac{m^2-1}{m^2+a} = \frac{1}{a+1}\sum_1^n \frac{\nu_s e_s^2}{m_s(p_s^2-p^2)},$$

where $a = \dfrac{1-\alpha}{\alpha}$. Now since the incident vibrations vary as $e^{\iota pt}$, and the natural vibrations as $e^{\iota p_1 t}$, $e^{\iota p_2 t}$, etc., they are respectively periodic in times $2\pi/p$, $2\pi/p_1$, $2\pi/p_2$, etc. If λ, λ_1, λ_2, etc., denote the corresponding wave-lengths measured in the free aether (not in the substance traversed by the waves)

$$\lambda = 2\pi c/p, \quad \lambda_1 = 2\pi c/p_1, \quad \lambda_2 = 2\pi c/p_2, \text{ etc.}$$

Thus
$$\frac{m^2-1}{m^2+a} = \frac{1}{4\pi^2c^2(1+a)}\sum_1^n \frac{\nu_s e_s^2\lambda^2\lambda_s^2}{m_s(\lambda^2-\lambda_s^2)}.$$

But
$$\frac{\lambda^2\lambda_s^2}{\lambda^2-\lambda_s^2} = \lambda_s^2 + \frac{\lambda_s^4}{\lambda^2-\lambda_s^2};$$

so that
$$\frac{m^2-1}{m^2+a} = \frac{1}{4\pi^2c^2(1+a)}\left\{\sum_1^n \frac{\nu_s e_s^2\lambda_s^2}{m_s} + \sum_1^n \frac{\nu_s e_s^2\lambda_s^4}{m_s(\lambda^2-\lambda_s^2)}\right\}.$$

Now when λ is infinite $\sum_1^n \dfrac{\nu_s e_s^2\lambda_s^4}{m_s(\lambda^2-\lambda_s^2)} = 0$, since every term in the summation vanishes. But when the wave-length is infinitely long and the period is infinitely slow, the case under consideration approaches continuously to that of an electrostatic field. The quantity κ, which enters into the equations of propagation, must therefore become identical with the dielectric coefficient as measured by electrostatic methods. Hence for infinitely long waves $m^2 = \kappa$, the dielectric coefficient of the medium. We therefore have

$$\frac{\kappa-1}{\kappa+a} = \frac{1}{4\pi^2c^2(1+a)}\sum_1^n \frac{\nu_s e_s^2\lambda_s^2}{m_s} \quad \ldots\ldots\ldots\ldots\ldots(7),$$

and $$\frac{m^2-1}{m^2+a}=\frac{\kappa-1}{\kappa+a}+\frac{1}{4\pi^2c^2(1+a)}\sum_1^n\frac{\nu_s e_s^2\lambda_s^4}{m_s(\lambda^2-\lambda_s^2)} \quad \text{......(8)}.$$

ν_s, e_s, λ_s and m_s are constants characteristic of the material under consideration, so that (8) may be written in the form

$$\frac{m^2-1}{m^2+a}=\frac{\kappa-1}{\kappa+a}+\frac{C_1}{\lambda^2-\lambda_1^2}+\frac{C_2}{\lambda^2-\lambda_2^2}+\ldots+\frac{C_n}{\lambda^2-\lambda_n^2} \quad \text{...(9)},$$

where C_1, C_2, ... C_n are constant quantities characteristic of the substance. There will be one of the terms C_1, C_2, etc., corresponding to each of the natural periods.

In the case of transparent colourless substances the natural periods must be either in the infra-red or in the ultra-violet part of the spectrum. If they are in the infra-red the synchronous wave-lengths λ_r will be large compared with λ, and if they are in the ultra-violet the λ_v's will be small compared with λ. Thus, as an approximation, we may write (9) in the form

$$\frac{m^2-1}{m^2+a}=\frac{\kappa-1}{\kappa+a}-\Sigma\frac{C_r}{\lambda_r^2}+\Sigma\frac{C_v}{\lambda^2}$$

$$=\text{Constant}+\Sigma\frac{C_v}{\lambda^2}, \text{ approx.} \ldots\ldots\ldots\ldots(9\,a).$$

In these formulae we have neglected the fractions $\frac{\lambda_v^2}{\lambda^2}$ and $\frac{\lambda^2}{\lambda_r^2}$ compared with unity. Since the refractive index m changes roughly in the same way as $\frac{m^2-1}{m^2+a}$, it follows from $(9a)$ that for these substances the refractive index will increase continuously as λ diminishes, in this region. Since the transparent colourless substances were the first to have their dispersion investigated, this type of dispersion is said to be normal. As we shall see, it can only be said to be normal provided we are a long way from the natural frequencies of the substances.

The behaviour of formulae (8) and (9) in the case of light whose frequency is close to that of the natural vibrations of the substance is most interesting. When $p=p_s$ for example, then $\lambda=\lambda_s$, and the corresponding term in (9) becomes infinite. If λ is slightly less than λ_s, then $C_s/(\lambda^2-\lambda_s^2)$ has a large negative value; and if λ is slightly greater than λ_s, then $C_s/(\lambda^2-\lambda_s^2)$ has a large positive value. As λ approaches λ_s from smaller values of

λ, $\frac{m^2-1}{m^2+a}$ approaches $-\infty$, and as λ approaches λ_s from larger values of λ, $\frac{m^2-1}{m^2+a}$ approaches $+\infty$. There is thus a sudden change in the value of $\frac{m^2-1}{m^2+a}$ from $-\infty$ to $+\infty$ at each of the synchronous frequencies. If we plot $\frac{m^2-1}{m^2+a}$ as ordinates against λ as abscissae the curve will have two branches, one extending to minus infinity, and the other returning from plus infinity at each of the values $\lambda=\lambda_1$, $\lambda=\lambda_2$, etc. Thus in crossing an absorption band in the direction of increasing wave-length, the refractive index will suddenly increase, and the rays close to the band on the red side will be more deviated than those on the violet. This kind of dispersion is said to be anomalous. It was discovered by Leroux* in experimenting with a prism filled with iodine vapour. The later investigations of Kundt† showed that it was related to the presence of absorption bands in the way indicated by the type of theory now under consideration. We also observe that if λ is less than every one of the values $\lambda_1, \lambda_2, \dots \lambda_n$, all of the terms $C_1/(\lambda^2-\lambda_1^2)$, etc., are of the same sign. Moreover in all the cases known at present $1+a$ is positive, and all the other quantities which make up the C's are essentially positive, so that each of the terms $C_1/(\lambda^2-\lambda_1^2)$, etc. is negative when λ is less than each of the values λ_1, λ_2, etc. Also if λ is greater than each of the quantities $\lambda_1, \lambda_2, \dots \lambda_n$, all of the quantities $C_1/(\lambda^2-\lambda_1^2), \dots C_n/(\lambda^2-\lambda_n^2)$ are positive. It follows that for sufficiently short waves m^2 is always less than κ, and for sufficiently long waves m^2 is always greater than κ. In the case of the latter assertion it is necessary to make one reservation, since m^2 clearly approaches the value $m^2=\kappa$ when λ becomes infinite. On the other hand, it follows from equation (5) that as λ becomes zero and p becomes infinite m^2 approaches the value unity. This is in accordance with experimental results, since the Roentgen Rays, which, as we shall see, may be looked upon as electromagnetic waves of very high frequency, are not deviated in passing through a prism.

A great many dispersion formulae, that is to say, formulae which connect the refractive index with the wave-length or frequency of the transmitted light, have been in vogue from time to time. Cauchy's formula

---

* *C. R.* vol. LV. p. 126 (1862).

† *Ann. der Phys.* vol. CXLII. (1871), and later papers in the same journal.

$$m = A_1 + A_2\lambda^{-2} + A_3\lambda^{-4} + \ldots$$

was the first to have a theoretical justification. Cauchy developed it from general considerations of wave theory on the hypothesis that the distance between the vibrating particles of the medium could not be regarded as completely negligible when compared with the wave-length. The formula thus obtained gives a fairly satisfactory representation, when three terms are used, of the dispersion of a number of substances. As it makes the refractive index increase continuously with diminishing wave-lengths (the constants are all positive), it fails absolutely to account for anomalous dispersion.

The idea that time rather than length was the determining common factor of the light and the matter which gave rise to dispersion, appears to have occurred first to Maxwell*. Maxwell supposed that when the atoms (we should now say electrons) were displaced, forces of restitution were called into play, and that there was also a resistance to their motion. A similar idea occurred somewhat later to Sellmeier, after whom the formula to which this theory gives rise is usually named. In the case of a single mode of vibration and in the absence of friction, the case contemplated by Sellmeier, the formula is

$$m^2 = A + \frac{B}{\lambda^2 - \lambda_s^2},$$

where λ is the wave-length of the light, and λ_s that corresponding to the natural vibration of the substance. In the case in which α (p. 73) $= 0$, formula (5) becomes

$$\begin{aligned} m^2 &= 1 + \sum_1^n \frac{\nu_s e_s^2}{m_s(p_s^2 - p^2)} \\ &= 1 + \sum_1^n \frac{\nu_s e_s^2 \lambda_s^2}{4\pi^2 c^2 m_s^2} + \sum_1^n \frac{\nu_s e_s^2 \lambda_s^4}{4\pi^2 c^2 m_s^2(\lambda^2 - \lambda_s^2)} \quad \ldots\ldots(10), \end{aligned}$$

so that Sellmeier's formula can be regarded as the particular case of formula (5) which arises when we put $\alpha = 0$ and $n = 1$.

It is still an open question whether the best dispersion formula is the Lorentz formula (5), or a generalized Sellmeier formula (see p. 176). It is however pretty clear that formula (10) is unsatisfactory, since, other things being equal, it makes m^2-1 vary as the density of the medium, and this is a great deal further from the truth than $\frac{m^2-1}{m^2+a} \propto \rho$. Both (5) and (10) give the same general kind of behaviour of m^2 when any one of the zeros $\lambda=\lambda_1$, $\lambda=\lambda_2$, etc., is crossed. The infinities of m^2 are however in different positions in the two cases. Thus in a formula of type (10), viz.

---

* Lord Rayleigh's *Coll. Papers*, vol. IV. p. 413.

$$m^2 = A_0 + \frac{A_1}{\lambda^2-\lambda_1^2} + \frac{A_2}{\lambda^2-\lambda_2^2} + \dots,$$

in the neighbourhood of, let us say $\lambda=\lambda_s$, we may put

$$m^2 = q_s + \frac{A_s}{\lambda^2-\lambda_s^2},$$

the rate of variation of the terms not involving λ_s being comparatively negligible. We observe that m^2 is negative from $\lambda=\lambda_s$ to $\lambda=\sqrt{\lambda_s^2-\frac{A_s}{q_s}}=\lambda_s\left(1-\frac{1}{2}\frac{A_s}{\lambda_s^2 q_s}\right)$ approx. Thus there is a range of wave-length $\delta\lambda_s = A_s/2q_s\lambda_s$ for which m is imaginary. In the case of a formula of type (5) we may write in the neighbourhood of an absorption band

$$\frac{m^2-1}{m^2+a} = q + \frac{A}{\lambda^2-\lambda_s^2},$$

or

$$m^2 = \frac{(1+aq)(\lambda^2-\lambda_s^2)+Aa}{(1-q)(\lambda^2-\lambda_s^2)-A}$$

$$= \frac{1+aq}{1-q} + A\,\frac{1+a}{(1-q)^2}\Big/(\lambda^2-\lambda_s'^2) \qquad \text{..............(11)},$$

where

$$\lambda_s'^2 = \lambda_s^2 + A/(1-q).$$

It is clear from (11) that m^2 changes from $+\infty$ to $-\infty$ as λ crosses from $\lambda>\lambda_s'$ to $\lambda<\lambda_s'$. m^2 then becomes continuously smaller numerically, but remains negative as to sign until it reaches the value zero, which is determined by

$$1 + aq + \frac{A(1+a)}{(1-q)(\lambda^2-\lambda_s'^2)} = 0,$$

or $\lambda^2 = \lambda_s^2 - Aa/(1+aq)$. Thus m is imaginary from

$$\lambda = \{\lambda_s^2 + A/(1-q)\}^{\frac{1}{2}} \text{ to } \lambda = \{\lambda_s^2 - Aa/(1+aq)\}^{\frac{1}{2}}.$$

Except in the immediate neighbourhood of an absorption band the value of the refractive index is practically the same whether we assume a dissipation term or not. Moreover the *general* character of the behaviour in the neighbourhood of an absorption band, as outlined above, is the same in both cases. It is important to bear in mind that the quantitative results are not the same, and many of the foregoing formulae are therefore to be taken as illustrative rather than representative. The same caution is to be observed in regard to the following treatment of residual rays, which neglects absorption. The more complete theory will be considered later.

The Residual Rays.

It is now necessary to interpret the imaginary value of m which, as we have seen, arises in a certain range of wave-lengths. If m is the refractive index of the material, the vectors in a plane wave propagated along the axis of z may be represented by

$$\epsilon^{-\iota\frac{2\pi}{\lambda}(ct-mz)}.$$

If m is imaginary, let us suppose it to be equal to $+\iota\beta$ where β is a real quantity. The expression now becomes

$$\epsilon^{-2\pi\frac{\beta}{\lambda}z}\,\epsilon^{-\iota\frac{2\pi}{\lambda}ct}.$$

If β is positive this represents a disturbance periodic in time which falls off indefinitely in amplitude as z increases. If β is negative it increases indefinitely. As the latter case would require infinite energy to maintain it, it may be left out of consideration. As β varies from ∞ to zero and λ is small, β/λ will be large over most of the range, so that the obvious interpretation of our result is that *light of the wave-lengths for which the value of m is imaginary is incapable of entering the medium.* It is not merely that the light is absorbed by the medium when it gets inside, as is the case for example in most coloured liquids which show body colour. It is almost unable to enter the surface. In fact all the light of the wave-length under consideration is totally reflected from the surface of the substance, even when the incidence is normal. For light of this particular range of wave-length the body behaves like a perfect reflector.

Many of the aniline dyes exhibit phenomena of this character in the visible spectrum. In fact something similar arises whenever the medium possesses intense absorption; so that it will be more convenient to consider the case of the aniline dyes when the theory of absorbing media has been discussed. It is clear from the considerations which have been urged that it is not *necessary* to have absorption of the type which is accompanied by degradation of energy for the rays to be unable to traverse the medium. It is only necessary that the period of the light vibrations should agree with one of the natural periods of the medium.

The best examples of this type of phenomenon have been found in the behaviour, in the infra-red region of the spectrum, of a number of insulators which are quite transparent to light in the visible spectrum. The most conspicuous examples are quartz, rock-salt, sylvite and fluorite. If the reflecting power of quartz, for example, is examined, it is found to be small, for normal incidence, for all wave-lengths from the visible spectrum up to about

$$7{\cdot}6\,\mu\,(1\mu = 10^{-4}\,\text{cm.}).$$

It then begins to increase rapidly as λ increases, until at a wave-length in the neighbourhood of $8{\cdot}1\,\mu$ quartz is almost as good a reflector of radiation as a metal. This state of things continues up to about $9\,\mu$, when the reflecting power begins to diminish. The transparency varies in the opposite way to the reflecting power. In fact between 8·1 and $9\,\mu$ quartz is so opaque that Nichols* was unable to detect any radiation through a layer of it only 2·5 wave-lengths in thickness. The relation between the reflecting power and the transparency of quartz is exhibited in the accompanying diagram (Fig. 25), which represents the results of Nichols's experiments.

Since the rays which correspond to the natural periods of substances are incapable of entering them, and so are always almost totally reflected, we are furnished with a new means of investigating the optical periods of substances. This method, due to Rubens, consists in submitting a beam of radiation from some source, such as a Nernst glower, to a series of successive reflexions, at incidences as nearly normal as possible, from surfaces of the material under investigation. The rays which are obtained after a sufficient number of reflexions will consist only of those which correspond to the natural periods of the substance. These rays have been named by Rubens, *Reststrahlen or Residual Rays.*

---

* *Ann. der Phys.* vol. LX. p. 401 (1897).

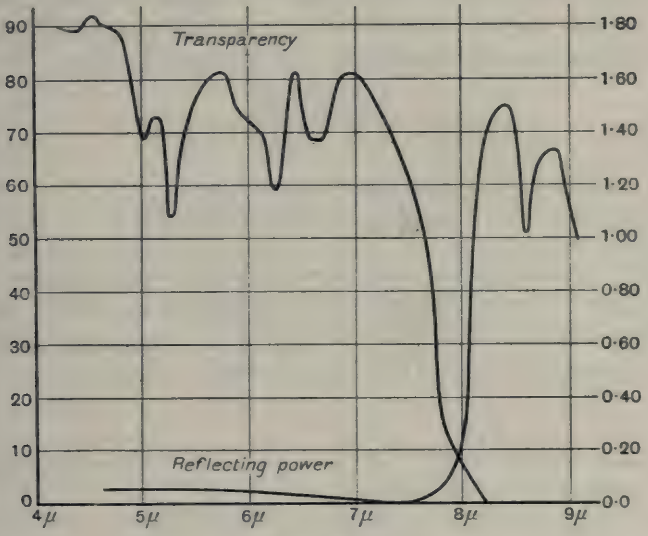


Fig. 25.

The wave-lengths of the residual rays for a number of substances are given in the following table:

Substance	Wave-length of Residual Rays in 10^{-4} cm.	Authority
Quartz (SiO_2)	8·1—9, 20·75	Nichols* Rubens and Nichols†
Fluorite (CaF_2)	24·4	Rubens and Nichols†
Sylvite (KCl)	62, 70	Rubens and Hollnagel‡
Rock-salt (NaCl)	45—48, 50—56	Rubens and Hollnagel‡
Wulfenite ($PbMoO_4$)	10·8—14	Coblentz§
Scheelite ($CaWO_4$)	10·8—13·2	Coblentz§
Corundum (Al_2O_3)	10·6—15	Coblentz§
Potassium Bromide (KBr)	76, 87	Rubens and Hollnagel‡

We have seen that according to the theories of dispersion with which we are concerned, the refractive index between certain limits of wave-length is an imaginary quantity. We shall consider the phenomenon of reflexion under these circumstances a little more fully, limiting ourselves to the comparatively simple case of normal incidence. Referring to Chapter VII, p. 132, we see that if the electric intensities in the incident, reflected and refracted waves are respectively X_1, X_3 and X_2, then

---

* *Loc. cit.*

† *Ann. der Phys.* vol. LX. p. 418 (1897).

‡ *Sitz. Preuss. Akad. Wiss.* p. 26 (1910).

§ *Jahr. der Rad.* vol. v. p. 1 (1908).

$$X_2 = \frac{2\sin\theta_2\cos\theta_1}{\sin\theta_2\cos\theta_1 + \sin\theta_1\cos\theta_2} X_1,$$

and

$$X_3 = \frac{\cos\theta_1\sin\theta_2 - \sin\theta_1\cos\theta_2}{\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2} X_1,$$

where θ_1 and θ_2 are the angles of incidence and refraction, and the magnetic permeabilities μ_1, μ_2 are taken as unity. In the case of normal incidence all the angles are vanishingly small, so that we may put $\sin\theta_1 = \theta_1 = m\sin\theta_2 = m\theta_2$. Hence

$$\left.\begin{aligned} X_2 &= \frac{2}{m+1} X_1 \\ \text{and} \quad X_3 &= -\frac{m-1}{m+1} X_1, \text{ when } \theta = 0 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(12).$$

When m is real the intensities of the light in the incident reflected and refracted waves are cX_1^2, cX_3^2 and mcX_2^2 respectively, and we observe that $mcX_2^2 + cX_3^2 = cX_1^2$, as is required by the principle of the conservation of energy. Calling the corresponding intensities I_1, I_3 and I_2 we have

$$\frac{I_2}{I_1} = \frac{4m}{(m+1)^2} \quad \text{and} \quad \frac{I_3}{I_1} = \left(\frac{m-1}{m+1}\right)^2.$$

Let us consider the change in the ratio of the reflected to the incident light as m travels over the range we are considering. When m is real and of the order unity, as in most cases of refraction in the laboratory, I_3 is of the same order as I_1 but is always less than I_1. As m increases to the value $+\infty$, I_3/I_1 becomes equal to unity. When m becomes zero, I_3 is still equal to I_1. Now consider what happens when m has the imaginary values which it takes between $\pm\infty$ and zero. We may put $m = i\beta$ where β is real. Then

$$X_3 = -\frac{m-1}{m+1} X_1 = +\frac{1-i\beta}{1+i\beta} X_1 = \frac{1-2i\beta-\beta^2}{1+\beta^2} X_1.$$

Let us put this into the form $\rho(\cos\theta + i\sin\theta) = \rho e^{i\theta}$, then ρ^2 will represent the ratio of the corresponding intensities, and θ will measure the difference of phase of the two beams under comparison. We have

$$\rho\cos\theta = +\frac{1-\beta^2}{1+\beta^2},\quad \rho\sin\theta = -\frac{2\beta}{1+\beta^2},$$

and
$$\rho^2 = \frac{(1-\beta^2)^2 + 4\beta^2}{(1+\beta^2)^2} = 1.$$

Thus the intensity of the reflected light is equal to that of the incident light for all the imaginary values of m which intervene between $+\infty$ and 0. It is clear that the intensity of the transmitted light must be zero throughout the same range of wavelength. This is at once evident for the limiting values $+\infty$ and 0, from the expression

$$\frac{I_2}{I_1} = \frac{4m}{(1+m)^2}.$$

The width of these regions appears to be quite considerable and is different according to the type of dispersion formula used*. For a formula of the Sellmeier type, we have seen (p. 156) that m is imaginary from $\lambda = \lambda_s$ to $\lambda = \sqrt{\lambda_s^2 - \frac{A_s}{q_s}}$. Thus the residual rays range from λ_s to $\lambda_s - A_s/2q_s\lambda_s$ approximately. Using the Lorentz type of formula (p. 156), m is imaginary from

$$\lambda = \sqrt{\lambda_s^2 + \frac{A}{1-q}} \text{ to } \lambda = \sqrt{\lambda_s^2 - \frac{Aa}{1+aq}}.$$

Thus the width of the residual rays is approximately

$$\delta\lambda = \frac{A(1+a)}{2\lambda_s(1-q)(1+aq)}.$$

We shall illustrate these results by considering the example furnished by rock-salt†. Paschen‡ has shown that the refractive index of rock-salt can be represented over the range from ·18μ to 22μ by the following dispersion formula of the Sellmeier type:

$$m^2 = b^2 + \frac{M_1}{\lambda^2 - \lambda_1^2} + \frac{M_2}{\lambda^2 - \lambda_2^2} + \frac{M_3}{\lambda^2 - \lambda_3^2},$$

---

* Havelock, *Roy. Soc. Proc.* A. vol. LXXXIV. p. 515 (1911).

† The numerical values are taken from the paper by Havelock (*loc. cit.*).

‡ *Ann. der Phys.* vol. XXVI. p. 130 (1908).

where
$$b^2 = 5{\cdot}680137,$$
$$M_1 = {\cdot}01278685, \qquad \lambda_1^2 = {\cdot}0148500\,\mu^2,$$
$$M_2 = {\cdot}005343924, \qquad \lambda_2^2 = {\cdot}02547414\,\mu^2,$$
$$M_3 = 12059{\cdot}95, \qquad \lambda_3^2 = 3600\,\mu^2.$$

[In the expressions above, and in the remaining formulae of this section, the values of the constants which are given are such that the values of the wave-lengths λ in terms of the unit $\mu = 1 \times 10^{-4}$ cm. have to be substituted.]

This gives two natural periods in the extreme ultra-violet and one in the infra-red at about $\lambda_3 = 60\mu$. In the immediate neighbourhood of the latter we find $m^2 = 5{\cdot}680142 + 12059{\cdot}95/(\lambda^2 - 3600)$. This gives for the lower limit of λ which corresponds to an imaginary value of m, the wave-length $38{\cdot}4\,\mu$. Thus the residual rays from rock-salt should extend from $60\,\mu$ to $38{\cdot}4\,\mu$; since the value $60\,\mu$ corresponds to the natural vibrations of that substance, according to the dispersion formula given by Paschen.

A dispersion formula of the Lorentz type has been given by Maclaurin*, which shows an even better agreement with the experimental values of the dispersion from ${\cdot}48\,\mu$ to $22\,\mu$. This is

$$\frac{m^2-1}{m^2+a} = \frac{\kappa-1}{\kappa+a} + \frac{C_1}{\lambda^2-\lambda_1^2} + \frac{C_2}{\lambda^2-\lambda_2^2},$$

where
$$a = 5{\cdot}51, \qquad \kappa = 5{\cdot}9,$$
$$C_1 = {\cdot}00191605, \qquad \lambda_1 = 0{\cdot}12652\,\mu,$$
$$C_2 = 683{\cdot}816, \qquad \lambda_2 = 51{\cdot}3\,\mu.$$

In the neighbourhood of λ_1, $q' = 0{\cdot}169652$, and the upper limit of λ which gives an imaginary refractive index is

$$\sqrt{\{\lambda_1^2 + C_1/(1-q')\}} = 0{\cdot}1353\,\mu.$$

The lower limit is $\sqrt{\{\lambda_1^2 - aC_1/(1+aq')\}} = 0{\cdot}1027\,\mu$. Thus in this region the residual rays would cover a range of ${\cdot}0226\,\mu$. In the infra-red the value of q_2' is $0{\cdot}429449$ and the upper limit is

$$\lambda = \sqrt{\{\lambda_2^2 + C_2/(1-q_2')\}} = 61{\cdot}9\,\mu.$$

The lower limit is

$$\lambda = \sqrt{\{\lambda_2^2 - aC_2/(1+aq_2')\}} = 38{\cdot}9\,\mu.$$

In this case total reflexion should occur over a range of $23\,\mu$.

---

* *Roy. Soc. Proc.* A. vol. LXXXI. p. 367 (1908).

The experimental results on the distribution of the energy among the different wave-lengths in the residual rays are not in accordance with the theoretical conclusions which we have just deduced. There is no evidence of the existence of finite stretches of spectrum in which the reflecting power is equal or nearly equal to unity. In every case the curves show a very sharp rise to a definite maximum, which is sometimes followed by a rather similar second maximum. The reflecting power at the maximum is generally comparable with 75 °/$_\circ$. There is no evidence of the existence of an extended region showing a reflecting power near to unity. It is probable, as we shall see, that this difference arises from the fact that we have omitted to consider the resistance term in the equations of motion, which corresponds to absorption.

Absorbing Media.

We have seen (p. 144) that absorption can be accounted for in a general way by the introduction of a fictitious retarding force proportional to the velocity of the vibrating particles. If the component, parallel to the axis of x, of the force of this type which acts on the sth particle is $-\beta_s\dot{x}_s$ the equations of motion become

$$m_s\frac{\partial^2 x_s}{\partial t^2}=e_s(E_x+\alpha_s P_x)-\frac{x_s}{\lambda_s}-\beta_s\frac{\partial x_s}{\partial t}\quad\ldots\ldots\ldots(13),$$

with similar equations for the displacements parallel to the axes of y and z. We shall suppose as before that the impressed electric intensity E is parallel to the axis of x. We shall leave out of account the displacements along the y and z axes, since in an isotropic medium, if E_y and E_z are equal to zero, the mean values, averaged in space, of y_s, z_s and their time derivatives will also be zero. Dropping the suffixes for the moment the equation of motion may be written

$$\left(mD^2+\beta D+\frac{1}{\lambda}\right)x=e(E_x+\alpha P_x).$$

If $E_x=E\epsilon^{ipt}$ and $P_x=P\epsilon^{ipt}$ the forced vibrations (see p. 147) will be given by

$$x=\left(mD^2+\beta D+\frac{1}{\lambda}\right)^{-1}e(E+\alpha P)\,\epsilon^{ipt}$$
$$=\frac{e(E+\alpha P)}{m(p_0^2-p^2)+i\beta p}\,\epsilon^{ipt}.$$

Thus the displacement of the sth electron is

$$x_s = \frac{e_s(E+\alpha P)}{m_s(p_s^2-p^2)+i\beta_s p}\,\epsilon^{ipt} \quad \text{..............(14).}$$

The polarization $P\epsilon^{ipt}$ is equal to $\sum_1^n \nu_s e_s x_s$, where ν_s is the number of electrons of type s per unit volume, and the summation is extended over all the electrons in the molecule. Proceeding in exactly the same way as before (p. 148), we find that the square of the refractive index is now given by

$$m^2 = 1 + \frac{\sum_1^n \dfrac{\nu_s e_s^2}{m_s(p_s^2-p^2)+i\beta_s p}}{1-\alpha\sum_1^n \dfrac{\nu_s e_s^2}{m_s(p_s^2-p^2)+i\beta_s p}},$$

or

$$\frac{m^2-1}{1+\alpha(m^2-1)} = \sum_1^n \frac{\nu_s e_s^2}{m_s(p_s^2-p^2)+i\beta_s p}$$

$$= \sum_1^n \frac{\nu_s e_s^2 m_s(p_s^2-p^2)}{m_s^2(p_s^2-p^2)^2+\beta_s^2p^2} - i\sum_1^n \frac{\nu_s e_s^2\beta_s p}{m_s^2(p_s^2-p^2)^2+\beta_s^2p^2} \quad \text{...(15).}$$

In all the important cases $\beta_s p$ is small compared with $m_s(p_s^2-p^2)$ except when p is very nearly equal to p_s. So that except in the immediate neighbourhood of the natural frequencies of the substance we shall have $\dfrac{m^2-1}{1+\alpha(m^2-1)} = \sum_1^n \dfrac{\nu_s e_s^2}{m_s(p_s^2-p^2)}$ very nearly. This is the same formula as we obtained previously on the assumption that there was no damping. Thus the introduction of a dissipative force has practically no influence on the phenomena save in the immediate neighbourhood of an absorption band.

In tracing the phenomena in the immediate neighbourhood of the natural frequencies, it is desirable to effect a simplification, otherwise the formulae become too cumbersome to manage. We shall suppose that in the immediate neighbourhood of p_s we can treat the contribution to the right-hand side of (15), which arises from electrons other than the sth one, as a constant quantity q independent of the frequency p. This will be legitimate if p varies over only a small range in the neighbourhood of p_s. It would probably lead to erroneous results if it were applied to a very broad absorption band. This difficulty, however, might be avoided if q were replaced by a few terms of an expansion in powers of

p, sufficient to take account of the normal dispersion caused by the more distant natural frequencies.

With this restriction then, we shall have in the neighbourhood of $p = p_0$, putting m_0 for the refractive index and dropping unnecessary suffixes

$$\frac{m_0^2 - 1}{1 + \alpha(m_0^2 - 1)} = q + \frac{\nu e^2}{m} \frac{1}{p_0^2 - p^2 + i\dfrac{\beta}{m}p} \quad \text{.........(16)}.$$

Thus $m_0^2 = \dfrac{\nu e^2(1-\alpha) + m(p_0^2 - p^2)(1 + q - \alpha q) + i\beta p(1 + q - \alpha q)}{m(p_0^2 - p^2)(1 - \alpha q) - \alpha\nu e^2 + i\beta p(1 - \alpha q)}$

and is complex. Let us put $m_0 = n(1 - i\kappa)$, so that

$$m_0^2 = n^2(1 - \kappa^2) - 2in^2\kappa.$$

Both n and κ are real. The vectors which specify a plane wave propagated along the axis of z will be proportional to $e^{i\frac{2\pi}{\lambda}(ct - m_0 z)}$ where λ is the wave-length measured in vacuo, c is the velocity of light in the free aether, and $2\pi c/\lambda = p$. Thus the vectors will be proportional to

$$e^{-\frac{2\pi n\kappa}{\lambda}z}\, e^{i\frac{2\pi}{\lambda}(ct - nz)} \quad \text{.....................(17)}.$$

As in the previous cases in which it has occurred, the real exponential represents a decay factor. The extinction coefficient is therefore equal to $\dfrac{2\pi}{\lambda}n\kappa$, and the medium exhibits absorption. The second factor shows that the phenomena are periodic in a distance equal to λ/n, so that the medium behaves otherwise as though its refractive index were equal to n. It is clearly of considerable importance to deduce the values of n and $n\kappa$ from equation (16).

After rationalizing the denominator in the expression for m_0^2 and then equating real and imaginary parts we find

$$n^2(1 - \kappa^2) = A + \frac{B\gamma}{\gamma^2 + \delta^2} \quad \text{.................(18)},$$

$$2n^2\kappa = \frac{B\delta}{\gamma^2 + \delta^2} \quad \text{......................(19)},$$

where

$$\left.\begin{aligned} A &= \frac{1 + q - \alpha q}{1 - \alpha q}, \quad B = \frac{\nu e^2}{(1 - \alpha q)^2} \\ \gamma &= m(p_0^2 - p^2) - \frac{\alpha\nu e^2}{1 - \alpha q}, \quad \delta = \beta p \end{aligned}\right\} \quad \text{.........(20)}.$$

Thus $$2n^2=\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}+A+\frac{B\gamma}{\gamma^2+\delta^2}\text{.........}(21),$$

$$2n^2\kappa^2=\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}-A-\frac{B\gamma}{\gamma^2+\delta^2}\text{.........}(22).$$

Feeble Absorption.

There is one particular case, which is of importance in nature, in which a considerable simplification of these formulae may be effected. A large number of substances exhibit absorption in varying degree throughout their volume, and are said to possess body colour. This is the case with most coloured solutions and minerals, and in fact with most coloured substances which are transparent to a considerable proportion of the visible spectrum. In these cases the absorption is small in a distance comparable with the wave-length of light, although it may be considerable in a distance comparable with 1 centimetre.

Consider the behaviour of waves for which p has a value such as to make $\gamma=0$. We see from p. 165 that the absorption causes the intensity of the light to fall off as $e^{-2\pi\frac{n\kappa}{\lambda}z}$. In a distance $z=\frac{\lambda}{2\pi}$ it will therefore diminish in the proportion of $e^{-n\kappa}$ to unity. Since the absorption in distances of this order is very small it follows that $e^{-n\kappa}$ must be very nearly equal to unity; so that $n\kappa$ must be a very small quantity. Turning to formula (22) we see that $n\kappa$ will only be small when $\gamma=0$, provided that B/δ is a small quantity. It follows that δ is a large quantity compared with B, and also that it is large compared with γ within a reasonable distance of the value of p corresponding to $\gamma=0$. In this region we may use, as an approximation, the formula

$$n\kappa=\frac{B\delta}{2A^{\frac{1}{2}}(\gamma^2+\delta^2)}\text{ ....}(23).$$

At the frequency for which the absorption is a maximum $\frac{\partial}{\partial p}(n\kappa)$ and therefore $\frac{\partial}{\partial p}\left(\frac{\delta}{\gamma^2+\delta^2}\right)$ vanishes. Hence

$$mp_0{}^2-\frac{\alpha ve^2}{1-\alpha q}-mp^2=\frac{\beta^2p^2}{mp_0{}^2-\frac{\alpha ve^2}{1-\alpha q}+3mp^2}.$$

By putting $p = p_0$ on the right-hand side as an approximation we see that the absorption maximum lies close to

$$p^2 = p_0^2 - \frac{1}{m}\frac{\alpha v e^2}{1 - \alpha q} - \frac{1}{m^2}\frac{\beta^2 p_0^2}{4p_0^2 - \frac{1}{m}\frac{\alpha v e^2}{1 - \alpha q}} \quad \text{.........(24).}$$

Near the natural frequencies of a gaseous substance we can put $A = 1$, $\alpha = 0$ and $q = 0$. In this case it may be shown* that satisfactory approximations for n and $n\kappa$ are

$$n = 1 + \frac{B\gamma}{2(\gamma^2 + \delta^2)} \quad \text{.......... ......(24 } a\text{),}$$

and

$$2n\kappa = \frac{B\delta}{\gamma^2 + \delta^2} \quad \text{........................(24 } b\text{).}$$

If we neglect the variation of δ compared with that of γ in (24 b), we see that the maximum absorption is given by $\gamma = 0$ or, to this order of approximation, by $p = p_0$. These relations have been utilized in interpreting certain spectroscopic and magneto-optical experiments. (Cf. Chap. XX.)

The Residual Rays from Absorbing Media.

If we are to discuss the phenomena which characterize the residual rays it is evidently necessary to consider the behaviour of the formulae for n and κ when B/δ is not small. Solving for κ we find

$$\kappa = \sqrt{\left\{1 + \left(\frac{A(\gamma^2 + \delta^2) + B\gamma}{B\delta}\right)^2\right\}} - \frac{A(\gamma^2 + \delta^2) + B\gamma}{B\delta} \quad \text{...(25),}$$

and, treating δ as constant

$$\frac{\partial\kappa}{\partial\gamma} = \frac{(2A\gamma + B)\left(\frac{A(\gamma^2 + \delta^2) + B\gamma}{B\delta}\right) - \sqrt{\left\{1 + \left(\frac{A(\gamma^2 + \delta^2) + B\gamma}{B\delta}\right)^2\right\}}}{B\delta\sqrt{\left\{1 + \left(\frac{A(\gamma^2 + \delta^2) + B\gamma}{B\delta}\right)^2\right\}}}.$$

This vanishes when $\gamma = -B/2A$ and when $\gamma = \pm\infty$. Treating n^2 and $n^2\kappa^2$ similarly we find that the maximum and minimum values of both n^2 and $n^2\kappa^2$ are roots of the cubic equation

$$4A\gamma^3 + 3B\gamma^2 - 4A\delta^2\gamma - B\delta^2 = 0.$$

---

* Cf. H. A. Lorentz, *Theory of Electrons*, pp. 154, 310.

Let us now turn our attention to the intensity of the light reflected at the surface of the medium under consideration. Confining ourselves to the case of normal incidence and following the method of Chapter VII, p. 138, we see that the ratio of the reflected to the incident electric intensity is

$$\frac{X_3}{X_1}=\frac{1-\epsilon}{1+\epsilon}, \text{ where } \epsilon=\frac{a_1\mu_1\cos\theta_2}{a_2\mu_1\cos\theta_1}=\frac{a_1\mu_1}{a_2\mu_2}.$$

For all the reflecting media that we have to deal with $\mu_1=\mu_2=1$, and since $a_1=c/\sqrt{\mu_1\kappa_1}$ and $a_2=c/\sqrt{\mu_2\kappa_2}$, we have

$$\epsilon=\frac{a_1}{a_2}=\sqrt{\frac{\kappa_2}{\kappa_1}}=m,$$

so that
$$\frac{X_3}{X_1}=-\frac{m-1}{m+1}=-\frac{n-i\kappa n-1}{n-i\kappa n+1}.$$

Putting this into the form $\rho e^{i\theta}$ we get

$$\rho\cos\theta=-\frac{n^2+n^2\kappa^2-1}{(n+1)^2+n^2\kappa^2},$$

and
$$\rho\sin\theta=+\frac{2n\kappa}{(n+1)^2+n^2\kappa^2},$$

so that
$$\rho^2=\frac{(n^2+n^2\kappa^2-1)^2+4n^2\kappa^2}{\{(n+1)^2+n^2\kappa^2\}^2} \qquad \text{..........(26)},$$

and
$$\tan\theta=-\frac{2n^2\kappa}{n^2+n^2\kappa^2-1} \qquad \text{....................(27)}.$$

As before, θ measures the change of phase on reflexion, and ρ^2 is the ratio of the intensity of the reflected and incident rays. The expressions which result on substituting from equations (21) and (22) are very complicated, and there does not appear to be any suitable approximation of a general character. The value of ρ^2 can only be obtained satisfactorily by numerical computation after the constants in the formulae have been determined. The character of the graph of ρ^2 as a function of p will be discussed after the theory of dispersion has been considered from a rather more general standpoint. In the same place we shall also review a number of the preceding results for absorbing media, using a rather simpler dispersion formula. (See p. 178.)

Generalized Theory of Dispersion.

In order properly to realize the limitations and approximations to which the foregoing theory of dispersion is subject, it is desirable to consider it briefly from a somewhat wider standpoint. From what has been said it is clear* that dispersion is essentially a dynamical problem in which the machinery is determined by the fundamental structure of the atom. Unfortunately we know very little about this structure, so that the necessary data cannot be stated very explicitly. Fortunately the methods of generalized dynamics enable us to find out a good deal about the behaviour of such a system even when we do not know much about its exact constitution.

In order to determine the motion it is necessary that certain functions of the state of the system should be known or, at least, be capable of definite specification. These are the Kinetic Energy, the Potential Energy, the Dissipation Function and the function which is equal to the work of the external forces. Let us consider these briefly in order.

Regarded as a dynamical system the optical medium consists of a system of electrons which may be treated as point charges subject, when undisturbed, to unknown conditions of equilibrium. The equilibrium is not necessarily a static one but may involve motion in orbits.

In any event the expression for the kinetic energy is quite simple, since it is equal to the sum of the energies of the individual electrons. If there are n electrons in any sufficiently large element of volume of the medium the kinetic energy δT which belongs to this element is

$$\delta T = \frac{1}{2}\sum_{s=1}^{s=n} m_s(\dot{x}_s^2 + \dot{y}_s^2 + \dot{z}_s^2) \quad \ldots\ldots\ldots\ldots\ldots(28),$$

where m_s is the mass and $\dot{x}_s$, $\dot{y}_s$ and $\dot{z}_s$ denote the components of the velocity of the sth electron.

To obtain an expression for the potential energy is much more complicated. If the equilibrium state involves steady motion, the potential energy will involve the velocity as well as the space coordinates. This makes the analysis considerably more complicated. As the present discussion is intended to be purely illustrative we shall forgo generality to the extent of supposing the undisturbed equilibrium to be a static one or at least one which can be treated as though it were static on the average. The effect of orbital motions has been considered to some extent by Larmor*.

---

* This is only true provided dynamics is adequate completely to account for the behaviour of atoms in this respect. This point is now doubtful, but, at least, it is of interest to examine the results to which generalized dynamics leads.

Under these conditions the most important term in the expression for the potential energy of any particular electron (suffix s) will undoubtedly be of the form

$$\alpha\,(x_s^2 + y_s^2 + z_s^2),$$

where α is a constant and x_s, y_s and z_s are the components of the displacement of the electron from the equilibrium position. This term represents the work done by the force of restitution contemplated by the former theory. If all the electrons always moved similarly and preserved similar geometrical relations in all the motions contemplated, the whole of the potential energy could probably be represented by a single constitutive term of this kind. But even when forced oscillations alone are considered, the relative displacement of the different kinds of electrons is affected by the frequency of the vibrations, so that it is necessary to consider the influence of the separate electrons on each other. Now each electron lies in the field of force of the doublets which are equivalent to the displacements of the other electrons, so that the complete expression for the potential energy of any particular electron will contain terms depending on the displacements of all the other electrons which lie within some considerable distance of it. Consider the part of the potential energy of the sth electron which arises from the displacement of some other specified electron which we may denote by the suffix a. Let the line joining e_a to the undisplaced position of e_s be r, where

$$r^2 = x^2 + y^2 + z^2.$$

When e_s is displaced an amount x_s, r will become

$$r' = \{(x + x_s)^2 + y^2 + z^2\}^{\frac{1}{2}}.$$

---

* *Phil. Trans.* A. vol. cxc. p. 236 (1897).

The potential energy of e_s in its undisplaced position, due to the doublet arising from e_a the moment of which is $e_a x_a$, is

$$e_s e_a x_a \frac{\partial}{\partial x_a}\left(\frac{1}{r}\right) = - e_s e_a x_a \frac{x}{r^3}.$$

The corresponding quantity for the displaced position of e_s is

$$e_s e_a x_a \frac{\partial}{\partial x_a}\left(\frac{1}{r'}\right) = - e_s e_a x_a \frac{x + x_s}{r'^3}.$$

There will therefore be a term in the potential energy of e_s, due to any other electron, which is of the form $e_s e_a x_s x_a f(r, \theta)$ where $f(r, \theta)$ depends upon the undisplaced positions of the electrons in space.

Similar remarks apply to the y and z components of the displacements of the electrons, so that the complete expression for the potential energy of the element of volume may be written

$$\delta W = \sum_{r=1}^{r=n} \sum_{s=1}^{s=n} \{A_{rs} x_r x_s + B_{rs} y_r y_s + C_{rs} z_r z_s + F_{rs} x_r y_s + G_{rs} x_r z_s + H_{rs} y_r z_s\} \ldots (29),$$

where the coefficients A, B, C, F, G, H involve the fundamental structure of the medium, but are independent of the displacements $x_r, y_r \ldots z_s$.

The nature of this expression for the potential energy calls for a little fuller consideration. The cross coefficients A_{rs} etc. are proportional to the inverse cube of the mutual distance of the electrons involved, so that they are small except for pairs of electrons which are quite near one another. On the other hand the number of electrons at a distance between r and $r + dr$ varies as $r^2 dr$; but, on account of the periodic character of the phenomena, the distant electrons are in layers which exert opposite and approximately equal effects. Thus the contribution to the potential energy of a particular electron, which arises from the cross terms, will come almost entirely from other electrons in its immediate neighbourhood; in other words the whole of this potential energy may be considered to arise from local causes. It is therefore legitimate to express it as a summation over the element of volume if this is taken fairly large. Unfortunately the size of the appropriate element will depend to some extent on the period of the vibrations, so this process can only be regarded as an approximation after all.

The summation in (29) can be split into two parts, one involving electrons in the same molecule and the other representing the influence of the electrons in the external molecules. The effect of the last named part is represented by the term αP in equations (1). This part of the summation will clearly vary with changes in the density of the medium. The effect of the part of the summation which depends on the other electrons in the same molecule is not considered in equations (1). It is not likely to be influenced much, if at all, by changes in the density of the medium.

The discussion on p. 143 shows that we do not know much about the nature of the forces leading to dissipation of energy in systems of this kind. It is therefore desirable to make some fairly general assumption about it and we shall suppose that the dissipation function δF is a quadratic function of the velocities of the electrons. The consequences of such a supposition have not, as yet, been shown to be incompatible with the results of experiments.

An expression for the work function δU may be found by considering the energy of a dielectric medium in which an electric field resides. If the electric intensity is E the energy per unit volume is $\frac{1}{2}DE = \frac{1}{2}E(E+P)$. The term $\frac{1}{2}E^2$ can be interpreted as the energy per unit volume of the space occupied by the dielectric, leaving $\frac{1}{2}EP$ as the work done by the field on the electrons. The work function is therefore

$$\delta U = \sum_{r=1}^{r=n} e_r \{E_x x_r + E_y y_r + E_z z_r\} \text{ ....(30).}$$

If we consider only plane polarized waves in which the electric intensity is parallel to the axis of x we can put

$$E_x = X, \quad E_y = E_z = 0$$

and

$$\delta U = \sum_{r=1}^{r=n} e_r X x_r \text{ ....(31).}$$

The equations of motion become greatly simplified when the functions T, W and F (dropping the δ's) become sums of squares. It is well known that by means of a linear transformation of coordinates any two of the functions T, W and F can be transformed into sums of squares of the new variables each multiplied by an appropriate coefficient. But we are unable to do this simultaneously for all three even when two of them are already sums of squares, as is the case now with T and possibly also F. For instance suppose we replace the $3n$ variables $x_1, y_1 \ldots z_n$ by $3n$ new variables, $q_1, q_2 \ldots q_{3n}$ which are linear functions of $x_1, y_1 \ldots z_n$. Let

$$\left.\begin{aligned} x_r &= a_r^1 q_1 + a_r^2 q_2 + a_r^3 q_3 + \ldots + a_r^{3n} q_{3n} \\ y_r &= b_r^1 q_1 + b_r^2 q_2 + b_r^3 q_3 + \ldots + b_r^{3n} q_{3n} \\ z_r &= c_r^1 q_1 + c_r^2 q_2 + c_r^3 q_3 + \ldots + c_r^{3n} q_{3n} \end{aligned}\right\} \ldots\ldots\ldots(32),$$

where the $(3n)^2$ coefficients $a_1^1 \ldots c_n^{3n}$ are constant quantities. The condition that this transformation should reduce the expression for T to a sum of squares is that the $\frac{3n(3n-1)}{2}$ equations between the $9n^2$ quantities a_1^1, a_1^2 etc., which arise when the coefficients of the cross terms in the new expression for T are equated to zero, should be satisfied. In a similar manner the functions W and F will lead to $\frac{3n(3n-1)}{2}$ equations each. As soon as we have more than two functions to reduce, the number of equations exceeds the number of variables, and the transformations will only be reducible if a number of the equations happen, on account of relations inherent in the original coefficients, to become identical. In the present case where we have three functions, $\frac{9n^2}{2} - \frac{9n}{2}$ of the equations would have to be alike. There is no reason to expect that this condition can be accurately satisfied with the kind of systems under discussion, so that we really ought to consider the general case in which T, W and F are any homogeneous quadratic functions. We shall do this very briefly later (p. 177).

Nevertheless there is one case of simultaneous reduction which is well worth considering,—that in which the kinetic and potential energies are converted into sums of squares by an appropriate linear transformation and F is assumed to be equal to a sum of squares of the velocity coordinates so obtained. There are two reasons why this treatment may be considered plausible. In the first place we are quite in the dark about the real nature of F, and in assuming it to be expressible as a sum of squares of the coordinates which enter into the normal forms of T and W we are really making the simplest possible assumption about it. Secondly there are a large number of cases in which the dissipation is small, and in these cases the assumption outlined will probably give a close approximation to the truth.

Introducing the variables $q_1, q_2 \dots q_{3n}$ which satisfy equations like (32) the kinetic energy becomes

$$T = \frac{1}{2} \sum_{s=1}^{s=3n} \beta_s \dot{q}_s^2 \quad \dots\dots\dots\dots\dots\dots\dots\dots (33),$$

where
$$\beta_s = \sum_{r=1}^{n} m_r \left(a_r^{s^2} + b_r^{s^2} + c_r^{s^2}\right),$$

provided that for all values of l and m between 1 and $3n$, excepting when $l = m$,

$$\sum_{r=1}^{n} m_r \left(a_r^l a_r^m + b_r^l b_r^m + c_r^l c_r^m\right) = 0.$$

There are $\frac{3n}{2}(3n-1)$ such relations between the $9n^2$ coefficients.

Turning to W we see that since

$$x_r x_s = \left(a_r^1 q_1 + a_r^2 q_2 + \dots + a_r^{3n} q_{3n}\right)\left(a_s^1 q_1 + a_s^2 q_2 + \dots + a_s^{3n} q_{3n}\right)$$

and

$$x_r y_s = \left(a_r^1 q_1 + a_r^2 q_2 + \dots + a_r^{3n} q_{3n}\right)\left(b_s^1 q_1 + b_s^2 q_2 + \dots + b_s^{3n} q_{3n}\right)$$

etc., we shall have

$$W = \frac{1}{2} \sum_{s=1}^{3n} \gamma_s q_s^2 \quad \dots\dots\dots\dots\dots\dots\dots\dots (34),$$

where

$$\tfrac{1}{2}\gamma_t = \sum_{r=1}^{n} \sum_{s=1}^{n} \left\{A_{rs} a_r^t a_s^t + B_{rs} b_r^t b_s^t + C_{rs} c_r^t c_s^t \right.$$
$$\left. + F_{rs} a_r^t b_s^t + G_{rs} a_r^t c_s^t + H_{rs} b_r^t c_s^t\right\}$$

provided that for all values of t and u, except $t = u$, between 1 and $3n$.

$$\sum_{r=1}^{n} \sum_{s=1}^{n} \left\{A_{rs}\left(a_r^t a_s^u + a_s^t a_r^u\right) + B_{rs}\left(b_r^t b_s^u + b_s^t b_r^u\right) + C_{rs}\left(c_r^t c_s^u + c_s^t c_r^u\right)\right.$$
$$\left. + F_{rs}\left(a_r^t b_s^u + a_s^t b_r^u\right) + G_{rs}\left(a_r^t c_s^u + a_s^t c_r^u\right) + H_{rs}\left(b_r^t c_s^u + b_s^t c_r^u\right)\right\} = 0.$$

There will again be $\frac{3n}{2}(3n-1)$ of these equations.

We also have

$$U = \sum_{r=1}^{n} X e_r x_r = \sum_{s=1}^{s=3n} \delta_s q_s \quad \dots\dots\dots\dots\dots\dots (35),$$

where
$$\delta_s = X \sum_{r=1}^{n} e_r a_r^s,$$

and we assume $$F=\frac{1}{2}\sum_{s=1}^{3n}\epsilon_s\dot{q}_s^2 \quad\text{.......................(36)},$$

where the ϵ's are constant coefficients.

The equations of motion are given for each q by the extended Lagrange's equation*

$$\frac{\partial}{\partial t}\left(\frac{\partial T}{\partial\dot{q}_s}\right)-\frac{\partial T}{\partial q_s}+\frac{\partial F}{\partial\dot{q}_s}+\frac{\partial V}{\partial q_s}=\frac{\partial U}{\partial q_s} \quad\text{..............(37)}.$$

They are therefore

$$\beta_s\ddot{q}_s+\gamma_s q_s+\epsilon_s\dot{q}_s=\delta_s \quad\text{..................(38)},$$

and if X varies as e^{ipt} the forced vibrations are given by

$$q_s=\frac{\delta_s}{-\beta_s p^2+\gamma_s+i\epsilon_s p} \quad\text{..................(39)}.$$

The natural vibrations will be obtained when the external electric intensity X is equal to zero. Those corresponding to the displacement q_s will therefore be determined by the equation

$$\beta_s\ddot{q}_s+\gamma_s q_s+\epsilon_s\dot{q}_s=0 \quad\text{..................(40)}.$$

They will be proportional to $e^{i\pi_s t}$ where π_s is a root of

$$-\beta_s\pi^2+i\epsilon_s\pi+\gamma_s=0.$$

Evidently π_s is complex and if we put $\pi_s\equiv p_s+ik_s$, p_s will be the frequency of the corresponding principal period and k_s its decay factor. Since

$$\pi_s^2=p_s^2-k_s^2+2ip_sk_s$$

we have $$-\beta_s(p_s^2-k_s^2)-\epsilon_s k_s+\gamma_s=0,$$

and $$\epsilon_s=2\beta_s k_s.$$

Hence $$p_s^2=\frac{\gamma_s}{\beta_s}-\frac{\epsilon_s^2}{4\beta_s^2} \quad\text{...... ....(41)},$$

and $$\gamma_s=\beta_s p_s^2+\frac{\epsilon_s^2}{4\beta_s}.$$

Thus $$q_s=\frac{\delta_s}{\beta_s(p_s^2-p^2)+\dfrac{\epsilon_s^2}{4\beta_s}+i\epsilon_s p}.$$

and $$x_r=\sum_{s=1}^{3n}a_r^s q_s=\sum_{s=1}^{3n}\frac{a_r^s\delta_s}{\beta_s(p_s^2-p^2)+\dfrac{\epsilon_s^2}{4\beta_s}+i\epsilon_s p} \quad\text{......(42)}.$$

---

* Lord Rayleigh, *Theory of Sound*, vol. I. chap. V.

The summations in the expressions above are to be extended over all the n electrons in any sufficiently large element of volume. If ν is the number of such elements in unit volume the polarization will be

$$\begin{aligned} P_x &= \sum_{r=1}^{n} \nu e_r x_r \\ &= \nu \sum_{r=1}^{n} e_r \sum_{s=1}^{3n} \frac{a_r^s \delta_s}{\beta_s (p_s^2 - p^2) + \frac{\epsilon_s^2}{4\beta_s} + i\epsilon_s p} \\ &= X \sum_{s=1}^{3n} \frac{\delta_s'}{\beta_s (p_s^2 - p^2) + \frac{\epsilon_s^2}{4\beta_s} + i\epsilon_s p} = (\kappa - 1) X \quad \ldots(43), \end{aligned}$$

where $$X\delta_s' = \nu\delta_s \sum_{r=1}^{n} e_r a_r^s.$$

Thus for a medium of unit magnetic permeability the complex refractive index m is given by

$$\begin{aligned} m^2 &= 1 + \sum_{s=1}^{3n} \frac{\delta_s'}{\beta_s (p_s^2 - p^2) + \frac{\epsilon_s^2}{4\beta_s} + i\epsilon_s p} \\ &= 1 + \sum_{s=1}^{3n} \frac{\Delta_s}{\psi_s^2 - p^2 + i\phi_s p} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(44), \end{aligned}$$

where

$$\left.\begin{aligned} \Delta_s &= \nu \left(\sum_{r=1}^{n} e_r a_r^s \sum_{r=1}^{n} e_r a_r^s \right) \div \sum_{r=1}^{n} m_r (a_r^{s2} + b_r^{s2} + c_r^{s2}) \\ \psi_s^2 &= p_s^2 + \left(\frac{\epsilon_s}{2} \div \sum_{r=1}^{n} m_r (a_r^{s2} + b_r^{s2} + c_r^{s2}) \right)^2 \\ \phi_s &= \epsilon_s \div \sum_{r=1}^{n} m_r (a_r^{s2} + b_r^{s2} + c_r^{s2}) \end{aligned}\right\} (45).$$

It is evident that the refractive index must be independent of n and ν except in the combination $n\nu$. The particular values of n and ν are arbitrary except that n has to be a sufficiently large number. The product $n\nu$ is equal to the number of electrons in unit volume of the substance and is therefore a characteristic constant. The requisite independence is secured by the fact that when n is large the constants which enter into (44) keep on repeating themselves for different values of s. Thus the summation in (44) is really a summation over the different principal modes of vibration in which each mode is multiplied by the number of times it occurs in unit volume. The total number of terms, coincident or otherwise, is equal to three times the number of electrons present in unit volume.

Formula (44) is of the Sellmeier type except for the inclusion of the dissipation term. With energy functions of the type now under consideration, the relation between the refractive index and the density of the substance is not an obvious one, since the constants Δ_s, ψ_s^2 and ϕ_s will involve the density in virtue of the relations on pp. 173—175. A formula of the Lorentz type would, however, arise if we assume that the only part of the force on an electron which depends on the density of the medium is $=\alpha P$, where α is a constant and P is the polarization. Formula (44) then becomes

$$\frac{m^2-1}{1-\alpha(m^2-1)}=\sum_{s=1}^{3n}\frac{\Delta_s'}{\psi_s'^2-p^2+2\phi_s'p}\text{............(46)},$$

where the constants are now somewhat different. Δ_s' is proportional to the density of the substance and ψ_s' and ϕ_s' are nearly independent of it.

In the general case in which the functions T, W and F are not simultaneously reducible to sums of squares the values of x_s etc. are the solutions of $3n$ simultaneous linear equations and can be written down in the form of determinants. Consequently these determinants enter into the expression for the refractive index and make it difficult to handle except by approximate methods. In general the symmetrical coefficients which lie along the axes of the determinants are large compared with the remaining unsymmetrical coefficients; so that the determinants can be expanded as a series of sums of products which decrease progressively in magnitude. In this way it can be shown that the Lorentz and Sellmeier types of formulae result in virtue of approximations which are equivalent to the physical assumptions which have already been made in deducing them.

We shall now return to the behaviour in the neighbourhood of an absorption band and the residual rays, using the simpler formula (44) instead of (15).

Absorption and Reflexion near the Critical Frequencies.

In the neighbourhood of one of the natural periods, say $p = p_s$, we shall have, as a sufficiently close approximation,

$$m^2 = q_s + \frac{\Delta_s}{\psi_s^2 - p^2 + i\phi_s p} \quad \text{.................} (47).$$

Putting $m = n(1 - \iota\kappa)$ where n and κ are real and positive we find, as on p. 165,

$$\left.\begin{aligned} n^2(1-\kappa^2) &= A + \frac{B\gamma}{\gamma^2+\delta^2} \\ 2n^2\kappa &= B\delta/(\gamma^2+\delta^2) \end{aligned}\right\} \quad \text{.................} (48),$$

where

$$\left.\begin{aligned} A &= q_s, & B &= \Delta_s \\ \gamma &= \psi_s^2 - p^2, & \delta &= \phi_s p \end{aligned}\right\} \quad \text{.................} (49),$$

$$\left.\begin{aligned} 2n^2 &= \sqrt{A^2 + \frac{B^2 + 2AB\gamma}{\gamma^2+\delta^2}} + A + \frac{B\gamma}{\gamma^2+\delta^2} \\ 2n^2\kappa^2 &= \sqrt{A^2 + \frac{B^2 + 2AB\gamma}{\gamma^2+\delta^2}} - A - \frac{B\gamma}{\gamma^2+\delta^2} \end{aligned}\right\} \quad \text{...} (50).$$

When the absorption is small in a distance compared with one wave-length we get, as before, to a first approximation

$$n\kappa = \frac{B\delta}{2A^{\frac{1}{2}}(\gamma^2+\delta^2)}.$$

Substituting the values of A, B, γ and δ we find $\frac{\partial(n\kappa)}{\partial p}$ vanishes if

$$\psi_s^2 - p^2 - \frac{\phi_s^2 p^2}{\psi_s^2 + 3p^2} = 0.$$

In the fraction we may put $\psi_s^2 = p^2$, as a sufficient approximation. Thus the value of p for which the absorption is a maximum is given by

$$p^2_{\text{max.}} = \psi_s^2 - \frac{\phi_s^2}{4} = p_s^2 \quad \text{....................} (51).$$

Thus the corresponding true natural frequency is the frequency for which the absorption is a maximum. It is somewhat less than the constant ψ_s^2 which enters into the dispersion formula.

We shall now turn to the problem of the intensity of the radiation reflected from a surface of the substance under consideration at normal incidence. Denoting this by ρ^2, considerations already brought forward show that it is given by

$$\rho^2 = \frac{(n^2+n^2\kappa^2-1)^2+4n^2\kappa^2}{\{(n+1)^2+n^2\kappa^2\}^2} = \frac{n^2(1+\kappa^2)-2n+1}{n^2(1+\kappa^2)+2n+1}.$$

Substituting the values of n and $n\kappa$ found previously, we get

$$\frac{\left(\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}-1\right)^2+2\left(\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}-A-\frac{B\gamma}{\gamma^2+\delta^2}\right)}{\left.\left\{\sqrt{\frac{1}{2}\left(\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}+A+\frac{B\gamma}{\gamma^2+\delta^2}\right)}+1\right\}^2+\frac{1}{2}\left(\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}}-A-\frac{B\gamma}{\gamma^2+\delta^2}\right)\right]^2}$$

$$\frac{\sigma^2+1-2\left(A+\frac{B\gamma}{\gamma^2+\delta^2}\right)}{\left.\sigma+1+2\sqrt{\frac{1}{2}\left(\sigma+A+\frac{B\gamma}{\gamma^2+\delta^2}\right)}\right\}^2} \quad \text{..(52)},$$

where $$\sigma=\sqrt{A^2+\frac{B^2+2AB\gamma}{\gamma^2+\delta^2}} \quad \text{..............(53)}.$$

We have seen that for substances which exhibit the phenomenon of body colour $n^2\kappa^2$ is small compared with unity, and for this to be the case for the particular value $\gamma=0$ it is necessary that δ should be large compared with B. There is no guarantee that this will be the case with substances which give rise to the residual rays, since Nichols has shown that in the case of quartz the amount of the residual rays which are transmitted through a slab of the substance only two to three wave-lengths thick is incapable of experimental detection. The value of the extinction coefficient, $n\kappa$, for such substances, may therefore be of the order unity or greater, and this corresponds to a value of B at least comparable with that of δ. It does not seem likely that there is any approximation of general application in the case of the residual rays which leads to any very marked simplification of the formulae. It is therefore necessary to evaluate the formulae in each particular case and this is a troublesome process. The constants A and B involved are obtainable from the corresponding constants in the usual Sellmeier dispersion formulae and so also is $\psi_s^2 (=\gamma+p^2)$. Determinations of δ from the experimental results do not seem to have been carried out as yet, but the value of δ is the most important factor in determining the maximum proportion of the incident energy reflected.

The precise nature of the curves which express ρ^2 as a function of γ or λ depends upon the values of the various constants. Nevertheless they always possess certain common features which are exhibited by the example in the accompanying figure. The ordinates represent the percentage of the incident energy which is reflected, i.e. they are the values of $100\,\rho^2$; and the abscissae are the wave-lengths ($\lambda = 2\pi c/p$) of the incident radiation. All the constants except ϕ_s have the same value in each of the graphs 1, 2 and 3. The common constants are:—$q_s = 2{\cdot}05$, $\Delta_s = 2{\cdot}563 \times 10^{28}$ and $\psi_s^2 = 4{\cdot}53 \times 10^{28}$. In graph 1, $\phi_s = 0$: in graph 2, $\phi_s = 1{\cdot}42 \times 10^{12}$: and in graph 3, $\phi_s = 1{\cdot}42 \times 10^{13}$. The vertical line at $\lambda = 8{\cdot}855$ indicates the value λ_s of λ which corresponds to the critical

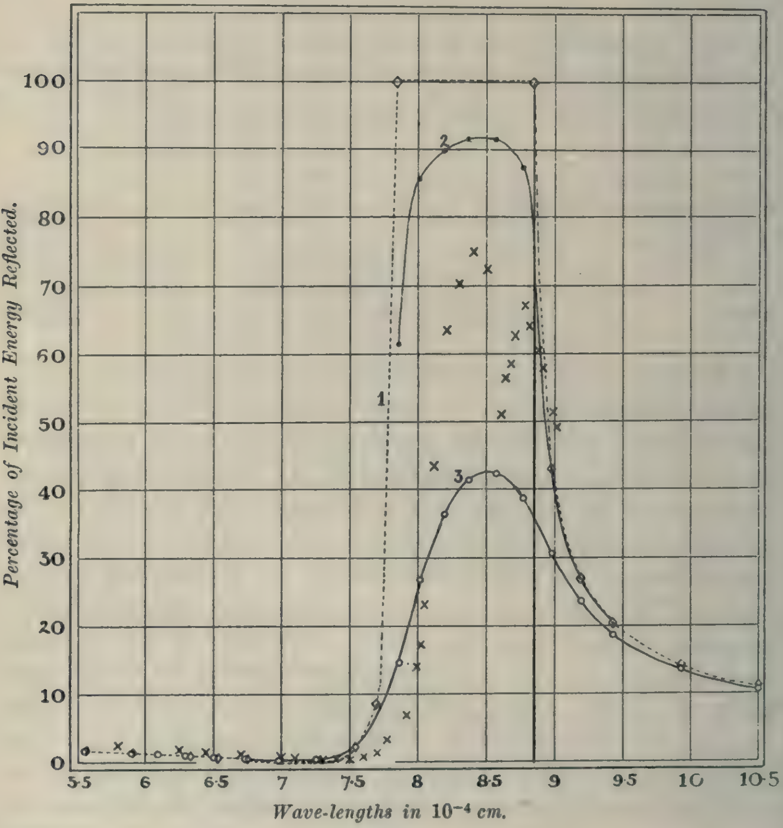


Fig. 26.

frequency ψ_s. The points marked thus: ×, are Nichols's experimental measurements of 100 ρ^2 for quartz. If we call λ_s the wave-length corresponding to ψ_s and to the value $\gamma=0$, λ_s will be equal to the constant wave-length whose square enters into the denominator in the usual Sellmeier dispersion formula. It is usually assumed that λ_s is identical with, or, at any rate, very close to the wave-length for which the energy is a maximum in the residual rays. It is evident from the figure that this assumption may be far from being justified.

Starting with values of λ which are less than λ_s (γ negative) ρ^2 has a small value which gradually decreases to a very small minimum. From this it rises very sharply to a maximum beyond which it again diminishes, but more slowly than it rose. Thus the curves are far from being symmetrical about the position for which ρ^2 is a maximum. The maximum value corresponds to a value of λ which is distinctly less than λ_s. The positions of the maximum and minimum may of course be obtained by differentiating the expression for ρ^2 with respect to γ, but the equation which results is of too high an order to be of much practical use. It happens, however, that the position of the minimum is very easily obtained with sufficient approximation, since it is practically coincident with the minimum value of the numerator in ρ^2. If we multiply the top and bottom by $\gamma^2+\delta^2$ the numerator remaining is

$$\{(A-1)\gamma+B\}^2+(A-1)^2\delta^2.$$

The minimum value of this is at $\gamma=-\frac{B}{A-1}$. The corresponding minimum value of ρ^2 is given very approximately, provided δ^2 is rather small, by

$$\rho^2=\frac{\delta^2}{4B^2}.$$

Thus the position of this minimum and the corresponding value of ρ^2 should give an important check on the constants in the dispersion formulae.

CHAPTER IX

THE FUNDAMENTAL EQUATIONS

THE fundamental equations of the electron theory may be regarded as a generalization or abstraction from the results of Chaps. III, V and VI. The electron theory assumes that matter is *nothing but* a distribution of electrified elements of volume in space. There are thus no magnetic charges in the sense in which there are ultimate electric charges or electrons. The magnetic fields which occur in nature arise entirely from the motion of the electrons. The simplest assumption which we can make as to the nature of the universal equations of the field is that they are identical with those which we have derived for the free aether containing electric charges. It is important to realize that this is an assumption, as it is sometimes regarded as self-evident. What we can be sure of is that the fundamental equations must degenerate into those for the free aether at points not in the immediate neighbourhood of material particles; but this is a very different thing from being sure that they are valid in the interior of an atom or an electron. The assumption of their universality is a hypothesis which will only be justified if the conclusions to which it leads are in agreement with deductions from experiments.

We therefore assume for the universal equations:

$$\operatorname{div} E = \rho \quad \text{...(1)},$$

$$\operatorname{div} H = 0 \quad \text{...(2)},$$

$$\operatorname{rot} E = -\frac{1}{c}\frac{\partial H}{\partial t} \quad \text{.................................(3)},$$

$$\operatorname{rot} H = \frac{1}{c}\left(\frac{\partial E}{\partial t} + \rho V\right) \quad \text{............................(4)},$$

where the mechanical force on an electric charge whose velocity is V relative to the measuring system is, per unit charge,

$$F = E + \frac{1}{c}[VH] \quad\text{.........................(5).}$$

It is necessary to show that these equations, which are associated with the name of Lorentz, are not inconsistent with any of our previous results. Looked at superficially they do appear to be inconsistent; since, by simply writing average values in equations (1) to (5) we do not arrive at equations which are obviously identical with those which we found to comprise the behaviour of dielectric and magnetic media in Chaps. III, V and VI. It is to be remembered, however, that the vectors defined as the electric and magnetic intensities and inductions respectively, in those chapters, were all average values of the true electric and magnetic intensities but formed in different ways. When this difference is taken into account the discrepancy will be seen to disappear. We shall now consider the equations in order from this point of view.

Equation (1) is supposed to apply to any element of volume however small. The corresponding equation $\operatorname{div} D = \rho$ is an equation between average values, and only applies to an element of volume which contains a very large number of electrons. In order to compare them let us integrate (1) over any sufficiently large volume. We have

$$\iiint \rho \, d\tau = \iiint \left\{ \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \right\} dx\,dy\,dz = \iint E_n \, dS,$$

where E_n represents the normal component of E at any point of the surface. But we have seen that the induction D_n is the average value of the force in the flat cavity perpendicular to D_n, so that the normal induction is nothing else than the average value of the intensity E taken over a surface perpendicular to it. This identification is only strictly true provided the surface is so large that the excess of polarization charges of a given sign inside of it is negligible. In other words, of the doublets whose axes are cut in two by the surface the difference between the number which leave their positive and those which leave their negative ends inside must be negligible. On the other hand, if the surface is not big enough to satisfy this condition, the meaning of the induction becomes indefinite.

In any event we saw at the end of Chap. III, that div $\bar{E} = \bar{\rho}$ always, provided that $\bar{\rho}$ represents the total average density of the charge whether it arises from conduction or polarization electrons or both. The apparently inconsistent equation div $D = \rho$ is only true provided the part of the volume density of the electrification which arises from the polarization electrons is left out of account.

Equation (15) of Chap. V, viz. div $B = 0$, is an equation over average values, and is consistent with div $H = 0$ for precisely the same reasons as those which establish the consistency of (1) and div $D = \rho$. In this case we do not need to consider the possibility of an excess of magnetic poles of a given sign being situated inside the surface. For as the elementary magnets consist of electric charges in motion, it is impossible to cut them in two in such a way as to separate the equivalent charges. The detailed formulation of the magnetic properties of bodies from this point of view will be left to a later chapter.

The equations obtained in Chapter VI also refer to average values of the dependent variables. The equation which is equivalent to (3) is

$$\text{rot } \bar{E} = -\frac{1}{c}\frac{\partial \bar{B}}{\partial t} \quad \dots\dots\dots\dots\dots\dots\dots(3a).$$

In order to show that these equations are consistent it is necessary to consider their geometrical interpretation. Each of them is an analytical expression of the fact that the line integral of the component, parallel to the contour, of the vector on the left, round any contour, is equal to the integral of the normal component of the vector on the right over any surface bounded by the same contour. Thus the $\bar{E}$ of $(3a)$ is the average value of the tangential component of the electric intensity taken round the contour. It is evidently equal to the average value of the E of (3) because $\bar{E}$ when derived from V as in Chapter IV is equal to the average value of E in a filamentous cavity. We have already seen that the average value of $\bar{B}_n$ over any surface is equal to the average value of H_n over the same surface. The equations (3) and $(3a)$ are therefore consistent with one another.

Quite similar considerations apply to (4) and equation (17) of Chap. VI, which we may write

$$\text{rot}\, \bar{H} = \frac{1}{c}\left(\frac{\partial \bar{D}}{\partial t} + \rho V\right) \quad \dots\dots\dots\dots\dots\dots(4a).$$

The average of the tangential component of H round a contour is evidently $\bar{H}$, and the average of E_n over the corresponding surface is clearly $\bar{D}_n$, so that these equations are also consistent. There is one point which is worth remarking in this case. In interpreting (4a) it is desirable so to choose the surface bounded by the contour that none of the electrons cross it during the interval under consideration. Otherwise there is a contribution to the ρV term owing to the motion of the electron. (4a) then becomes inconsistent with itself since the surface integral does not have the same value over all surfaces terminated by the same contour. This difficulty may be overcome either by choosing the surface so that the polarization electrons do not cross it, or by taking the element of time large enough to include the average value of effects arising from such translation. This is zero because the motion of an electron across the bounded surface is equivalent to the creation of a separate doublet with its like charge in the new position of the electron, and its unlike charge in the old position. The creation of this doublet introduces a local term in the force which just wipes out the effect of the motion of the charge across the boundary. These remarks are pertinent to equation (4a) only. Equation (4) is always consistent with itself, and is consistent with (4a) when the latter is self-consistent.

In comparing equation (5) with the corresponding equation

$$\bar{F} = \bar{E} + \frac{1}{c}[V \,.\, B] \dots\dots\dots\dots\dots\dots\dots(5a),$$

the agreement of the first term on the right is clear enough, but the second requires fuller consideration. Here we have to deal with the average value of H taken along a line to which H is normal. It is difficult to see how this may be done directly, but an indirect method may be employed. Considering the case where $E = 0$ (at least so far as average values are concerned) let us apply the universal equation $F = \frac{1}{c}[V \,.\, H]$ to find the force acting on any circuit carrying an electric current embedded in a material medium. If the strength of the current is i, (5) gives for the resultant force on the circuit the value $\frac{1}{c}\int[ids\,.\,H]$ taken round the circuit s. By an argument similar to that in Chap. V, p. 83, it follows that the force is equal to i/c times the rate of change of $\iint H_n dS$ over any surface having the same contour. The H_n in this integral is, of course, the normal component of the universal magnetic intensity. But, as we have seen,

$$\iint H_n dS = \iint B_n dS.$$

Thus it follows that (5) and ($5a$) are consistent and, incidentally, that the average value of H taken along a line normal to the direction of H is $\mu\bar{H} = B$. From this the analogy between the electric and magnetic vectors would lead us to expect that $\bar{E}_n = \frac{1}{s}\int E_n ds = \kappa\bar{E} = D$, the suffix n denoting that the vector is perpendicular to the direction of integration.

By dividing the space up by means of tubes of induction, it is clear that the average values of the universal expressions $\frac{1}{2}E^2$ and $\frac{1}{2}H^2$ for the electric and magnetic energy densities respectively are equal to $\frac{1}{2}\kappa\bar{E}^2$ and $\frac{1}{2}\mu\bar{H}^2$. This is only true provided we neglect constant terms which may be regarded as representing the intrinsic energy of the electrons and of the molecular magnets.

The Differential Equations satisfied by the Vectors when Charges are present.

In Chapter VII we were concerned with the solution of equations (1) to (4), and the extensions of them, which have just been considered, in the cases in which the density ρ of the charges was everywhere zero. The results thus obtained naturally applied to the propagation of electromagnetic effects in insulators, including the free aether as a particular case. We shall now consider the nature of the solutions in the more general case, when electric charges are present and contribute to the resulting phenomena by their motions and the forces they exert.

The new equations, analogous to $\nabla^2 E = \frac{\kappa\mu}{c^2}\frac{\partial^2 E}{\partial t^2}$, which are satisfied by E and H, may be obtained rather more easily than would otherwise be the case if we first prove the general theorem

$$\text{grad div } A - \nabla^2 A = \text{rot rot } A,$$

which is true if A is any vector point function. The components of rot A are

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z},$$

$$\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x},$$

$$\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}.$$

So that the x component, for example, of the rotation of the rotation of A is

$$\frac{\partial}{\partial y}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) - \frac{\partial}{\partial z}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) - \frac{\partial^2 A_x}{\partial y^2} - \frac{\partial^2 A_x}{\partial z^2}$$

$$= \frac{\partial}{\partial x}\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) - \left(\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2}\right).$$

Since a similar result follows for the other two components we have

$$\text{grad div } A - \nabla^2 A = \text{rot rot } A \quad \text{...............(6).}$$

In order to obtain the differential equation satisfied by E we differentiate equation (4) with respect to t and obtain

$$\frac{\partial^2 E}{\partial t^2} + \frac{\partial}{\partial t}(\rho V) = c \text{ rot } \frac{\partial H}{\partial t}.$$

Substituting the value of $\dot{H}$ from (3) we get

$$\frac{\partial^2 E}{\partial t^2} + \frac{\partial}{\partial t}(\rho V) = -c^2 \text{ rot rot } E,$$

whence, from (6) and (1),

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \text{grad } \rho + \frac{1}{c^2}\frac{\partial}{\partial t}(\rho V) \quad \text{............(7).}$$

In a similar manner, starting with (3) we find

$$\nabla^2 H - \frac{1}{c^2}\frac{\partial^2 H}{\partial t^2} = -\frac{1}{c}\text{ rot }(\rho V) \text{..................(8).}$$

Each of these equations is a vector equation and is equivalent to three separate equations between each of the three components. Thus if the components of E, H and V are $E_x, E_y, E_z, H_x, H_y, H_z, V_x, V_y, V_z$ respectively, the x components are given by the Cartesian equations

$$\nabla^2 E_x - \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2} = \frac{\partial \rho}{\partial x} + \frac{1}{c^2}\frac{\partial}{\partial t}(\rho V_x) \quad \text{.................(9)},$$

and

$$\nabla^2 H_x - \frac{1}{c^2}\frac{\partial^2 H_x}{\partial t^2} = -\frac{1}{c}\left\{\frac{\partial(\rho V_z)}{\partial y} - \frac{\partial(\rho V_y)}{\partial z}\right\} \quad \text{......(10)}.$$

There are four other similar equations for the other components.

The nature of the solutions of equations (9) and (10) may be discovered by considering the equation

$$\nabla^2 \psi - \frac{1}{c^2}\frac{\partial^2 \psi}{\partial t^2} = \omega \quad \text{.....................(11)},$$

where ω is a function of x, y, z and t. In the electrical problems which we shall have to consider ω is a given function of these variables. The solutions of (11) have a certain degree of resemblance to the potential in the theory of attractions. The potential V satisfies the equation $\nabla^2 V = \rho$, where ρ is the density of the attracting matter, measured in suitable units. As is well known the integral of this equation is

$$V = \frac{1}{4\pi}\iiint \frac{\rho}{r}\, d\tau \quad \text{.......................(12)}.$$

Thus the potential at any point P is obtained if we take the element, $\rho d\tau$, of mass at any point, divide by $4\pi r$ where r is the distance from P, and integrate throughout space. We shall see that a precisely similar result holds for the functions ψ which are the solutions of (11). The only difference lies in the fact that in calculating the values of ψ we replace ρ in (12), not by the instantaneous value of ω, the function on the right-hand side of (11), but by the value which this function had at the point of integration at an instant r/c previously, where r is the distance from the point at which E or H is to be calculated.

If in (11) we introduce a new independent variable $u = ict$ the equation becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{\partial^2 \psi}{\partial u^2} = \omega.$$

The left-hand side would be the value of $\nabla^2\psi$ in rectangular coordinates in a four-dimensional space; so that the problem of finding the solutions of (11) can be looked upon as the problem of finding the potential in a four-space. (11) is an example of a number of electromagnetic equations whose symmetry is improved when the time t is replaced by the imaginary variable $u = ict$.

Kirchhoff's Solution.

A very complete discussion of the solution of (11) was given in 1883 by Kirchhoff* in connection with the theory of the propagation of light. As a preliminary to solving (11) let us introduce an auxiliary function χ which satisfies the equation

$$\frac{\partial^2\chi}{\partial r^2} + \frac{2}{r}\frac{\partial\chi}{\partial r} = \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} \quad\text{......................(13).}$$

This is the equation to which (11) reduces when the right-hand side is put equal to zero and ψ is a function only of t and the distance r from a fixed point. If we put $\phi = r\chi$, (13) becomes

$$\frac{\partial^2\phi}{\partial r^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} \quad\text{......................(13 a).}$$

The most general solution of this equation is (see p. 117)

$$\phi = F(t + r/c) - F(t - r/c),$$

where F is any function whatever. The two terms correspond physically to disturbances propagated in opposite directions with velocity c. We shall only consider one of them and take

$$\phi = F(t + r/c)$$

giving
$$\chi = \frac{1}{r}F\left(t + \frac{r}{c}\right) \quad\text{......................(14),}$$

where F is a perfectly arbitrary function.

Next consider the integral

$$J = \iiint (\psi\nabla^2\chi - \chi\nabla^2\psi)\, d\tau \quad\text{................(15),}$$

---

* *Ann. der Phys. und Chemie*, vol. XVIII. p. 663 (1883).

taken throughout a closed volume limited by an internal surface σ and an external surface S. By Green's Theorem

$$J = -\iint \left(\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right) d\sigma - \iint \left(\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right) dS \quad \text{...(16)},$$

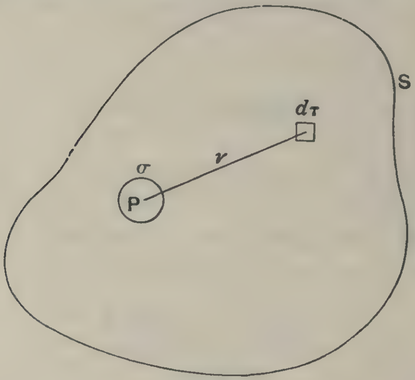


Fig. 27.

the normals being directed into the enclosed volume. We also have from (15), (11) and (13)

$$J = -\iiint \chi\omega\, d\tau + \frac{1}{c^2}\iiint \left(\psi \frac{\partial^2 \chi}{\partial t^2} - \chi \frac{\partial^2 \psi}{\partial t^2}\right) d\tau$$

$$= -\iiint \chi\omega\, d\tau + \frac{1}{c^2}\frac{\partial}{\partial t}\iiint \left(\psi \frac{\partial \chi}{\partial t} - \chi \frac{\partial \psi}{\partial t}\right) d\tau \quad \text{...........(17)}.$$

From (16) and (17)

$$-\iint \left(\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right) d\sigma = -\iiint \chi\omega\, d\tau + \iint \left\{\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right\} dS$$
$$+ \frac{1}{c^2}\frac{\partial}{\partial t}\iiint \left(\psi \frac{\partial \chi}{\partial t} - \chi \frac{\partial \psi}{\partial t}\right) d\tau.$$

This is true for all values of t. Let us integrate it with respect to dt between limits t_1 and t_2. We then get

$$-\int_{t_1}^{t_2} dt \iint \left(\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right) d\sigma$$
$$= -\int_{t_1}^{t_2} dt \iiint \chi\omega\, d\tau + \int_{t_1}^{t_2} dt \iint \left\{\psi \frac{\partial \chi}{\partial n} - \chi \frac{\partial \psi}{\partial n}\right\} dS$$
$$+ \frac{1}{c^2}\left|\iiint \left(\psi \frac{\partial \chi}{\partial t} - \chi \frac{\partial \psi}{\partial t}\right)\right|_{t_1}^{t_2} \quad \text{.......................(18)}.$$

Let us now return to the consideration of the function F. This may be any function of the argument $t + r/c$. We shall suppose that it is such a function that it takes the value zero for all values of the argument except those in the immediate neighbourhood of the particular value zero. We shall then have $F(x) = 0$ unless x lies between, let us say, $\pm\epsilon$ where ϵ is a very small quantity. We shall also suppose that

$$\int_{-\epsilon}^{+\epsilon} F(x)\,dx = 1 \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots(19).$$

Since $F(x)$ is zero unless x is between $\pm\epsilon$ we evidently also have

$$\int_{-\infty}^{\infty} F(x)\,dx = \int_{-\epsilon}^{+\epsilon} F(x)\,dx = 1.$$

If the value of r is fixed

$$\int_{t_1}^{t_2} F\left(t + \frac{r}{c}\right) dt = \int_{t_1 + r/c}^{t_2 + r/c} F(x)\,dx = 1$$

provided $t_2 + r/c > \epsilon$ and $t_1 + r/c < -\epsilon$. Moreover, if we make ϵ infinitesimal but still suppose F to have the property $\int_{-\epsilon}^{+\epsilon} F(x)\,dx = 1$ we shall have, if ω is any function of r and t,

$$\int_{t_1}^{t_2} \omega F\left(t + \frac{r}{c}\right) dt = \omega' \int_{t_1}^{t_2} F\left(t + \frac{r}{c}\right) dt = \omega' \quad\ldots\ldots\ldots(20),$$

where ω' is the value of ω at the instant $t = -r/c$. This follows since, except when t lies between $\pm\epsilon$, $F = 0$, and throughout this infinitesimal interval ω may be considered constant.

Now let t_2 have a definite positive value and t_1 a very large definite negative value, $-t_1$ being so large that for all points in the enclosed volume $t_1 + r/c < -\epsilon$. Then the values of χ which occur in the integrated part of (18) all vanish. So also do the values of $\frac{\partial\chi}{\partial t}$, since the derivatives of χ are also zero except between $\pm\epsilon$.

We may write the term containing ω in (18) in the form

$$-\iiint \frac{1}{r}\,d\tau \int_{t_1}^{t_2} \omega F\left(t + \frac{r}{c}\right) dt.$$

This is equal to $-\iiint \frac{\omega'}{r}\,d\tau$ where ω' has the same meaning as in (20). In a similar way

$$\int_{t_1}^{t_2} dt \iint \chi \frac{\partial\psi}{\partial n}\,dS = \iint \frac{1}{r}\left(\frac{\partial\psi}{\partial n}\right)' dS,$$

where $\left(\frac{\partial\psi}{\partial n}\right)'$ denotes the value of $\frac{\partial\psi}{\partial n}$ at the point of integration (distant r from P) at the instant $t=-r/c$. The corresponding integral with respect to $d\sigma$ may be similarly treated.

We also have

$$\frac{\partial\chi}{\partial n}=\frac{\partial r}{\partial n}\left\{\frac{1}{rc}F'\left(t+\frac{r}{c}\right)-\frac{1}{r^2}F\left(t+\frac{r}{c}\right)\right\}.$$

Thus
$$\int_{t_1}^{t_2}dt\iint\psi\frac{\partial\chi}{\partial n}dS$$

$$=\iint\frac{\partial r}{\partial n}dS\int_{t_1}^{t_2}dt\left\{\frac{1}{rc}\psi F'\left(t+\frac{r}{c}\right)-\frac{1}{r^2}\psi F\left(t+\frac{r}{c}\right)\right\}$$

$$=\iint\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\psi'\,dS+\iint\frac{1}{rc}\frac{\partial r}{\partial n}dS\int_{t_1}^{t_2}\psi F'\left(t+\frac{r}{c}\right)dt,$$

where ψ' is the value of ψ at the point of integration at the instant $t=-\frac{r}{c}$. The right-hand integral may be integrated by parts, giving

$$\iint\frac{1}{rc}\frac{\partial r}{\partial n}dS\left\{\left|\psi F\left(t+\frac{r}{c}\right)\right|_{t_1}^{t_2}-\int_{t_1}^{t_2}\frac{\partial\psi}{\partial t}F\left(t+\frac{r}{c}\right)dt\right\}$$

$$=-\iint\frac{1}{rc}\frac{\partial r}{\partial n}\left(\frac{\partial\psi}{\partial t}\right)'dS,$$

since $F\left(t+\frac{r}{c}\right)$ vanishes at the limits. Here $\left(\frac{\partial\psi}{\partial t}\right)'$ denotes the value of $\frac{\partial\psi}{\partial t}$ at the point of integration at the instant $t=-\frac{r}{c}$, in accordance with our former notation.

The left-hand side of (18) may be treated similarly, giving rise to

$$\iint\left\{\frac{1}{r}\left(\frac{\partial\psi}{\partial n}\right)'-\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\psi'+\frac{1}{cr}\frac{\partial r}{\partial n}\left(\frac{\partial\psi}{\partial t}\right)'\right\}d\sigma.$$

Now let the surface σ become coincident with a sphere of infinitesimal radius ρ about the point P. Then $\frac{\partial}{\partial n}\equiv\frac{\partial}{\partial\rho}$ and the left-hand side becomes

$$\iint\left\{\frac{1}{\rho}\left(\frac{\partial\psi}{\partial n}\right)'+\frac{1}{\rho^2}\psi'+\frac{1}{c\rho}\left(\frac{\partial\psi}{\partial t}\right)'\right\}\rho^2\sin\theta\,d\theta\,d\phi.$$

When ρ is made to become very small the terms in $\frac{1}{\rho}$ become infinitesimal compared with $\frac{\psi'}{\rho^2}$. So that the left-hand side becomes identical with $4\pi\Psi$, where Ψ is the value of ψ at the point P at the instant $t = 0$. Hence, we have

$$\Psi = -\frac{1}{4\pi}\iiint \frac{\omega'}{r}\,d\tau - \frac{1}{4\pi}\iint \left\{\frac{1}{r}\left(\frac{\partial\psi}{\partial n}\right)' - \frac{\partial}{\partial n}\left(\frac{1}{r}\right)\psi' + \frac{1}{cr}\frac{\partial r}{\partial n}\left(\frac{\partial\psi}{\partial t}\right)'\right\} dS \qquad \text{......(21).}$$

Now let the surface S recede to an infinite distance and suppose that at infinitely distant points the functions ψ, $\frac{\partial\psi}{\partial n}$, and $\frac{\partial\psi}{\partial t}$, all have the value zero until a definite time T, then when r becomes infinite the time $t = -\frac{r}{c}$, to which ψ', $\left(\frac{\partial\psi}{\partial n}\right)'$, and $\left(\frac{\partial\psi}{\partial t}\right)'$ in the surface integral refer, is always less than T, so that every element of the integral is zero. This supposition is legitimate physically since we always presuppose that physical phenomena are independent of past or present occurrences at an infinite distance. We thus see that the value Ψ at the point P at the time t is equal to

$$-\frac{1}{4\pi}\iiint \frac{\omega'}{r}\,d\tau \qquad \text{.......................(22),}$$

where the integral is taken throughout space, and the value of ω for each element of volume is that which it possessed at the instant $t - \frac{r}{c}$.

The Propagated Potentials.

The physical interpretation of the result we have just obtained is very simple. It means that the values of the electric and magnetic intensities at any particular point P at any instant are not, in general, determined by the state of the rest of the field at that particular instant, but by its previous history. The effects at P, in so far as they are due to a particular element of volume distant r from P, depend upon the state of that element of volume at a time r/c earlier than the instant considered. This time r/c is equal to the time which would be required for light to travel from the distant element to the point P. The nature of the field is therefore such as would arise if each portion of it were constantly emitting *disturbances* which were *propagated from it in all directions with the velocity of light.*

When we come to the actual calculation of the values of E and H in particular cases it is found that equations (7) and (8) are unsuitable owing to the values of ω, given by the right-hand sides of them, being somewhat complicated. The calculations may be simplified by the introduction of two new functions, the scalar potential ϕ and the vector potential U, from which E and H may afterwards be derived by appropriate operations. We shall now prove that

$$H = \text{rot } U \qquad \text{.......................(23)},$$

and

$$E = -\frac{1}{c}\frac{\partial U}{\partial t} - \text{grad } \phi \qquad \text{.....................(24)}$$

if ϕ and U satisfy the equations

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\rho \qquad \text{.......................(25)},$$

and

$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = -\frac{1}{c}\rho V \qquad \text{.................(26)}.$$

We shall prove first of all that a function U always exists such that $H = \text{rot } U$. This function is in fact

$$U = \frac{1}{4\pi}\iiint \frac{\text{rot } H}{r}\, d\tau \qquad \text{..................(27)}.$$

For if (27) is true we have

$$U_x = \frac{1}{4\pi}\iiint \frac{(\text{rot } H)_x}{r}\, d\tau = \frac{1}{4\pi}\iiint \frac{1}{r}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) d\tau.$$

Now the values of H in the integral refer to the different points of integration and not to the point at which U is measured. Let x, y, z be the coordinates of the point at which U is required and a, b, c the coordinates of the element of volume $d\tau$. Then $r^2 = (x-a)^2 + (y-b)^2 + (z-c)^2$ and the equation above may be written more clearly as

$$U_x = \frac{1}{4\pi}\iiint \left(\frac{\partial H_c}{\partial b} - \frac{\partial H_b}{\partial c}\right) \frac{da\,db\,dc}{\sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2}},$$

and $$(\text{rot } U)_x = \frac{\partial U_z}{\partial y} - \frac{\partial U_y}{\partial z}$$

$$= \frac{1}{4\pi}\left\{\frac{\partial}{\partial y}\iiint\left(\frac{\partial H_b}{\partial a} - \frac{\partial H_a}{\partial b}\right)\frac{d\tau}{r} - \frac{\partial}{\partial z}\iiint\left(\frac{\partial H_a}{\partial c} - \frac{\partial H_c}{\partial a}\right)\frac{d\tau}{r}\right\}$$

$$= \frac{1}{4\pi}\left\{\iiint\left(\frac{\partial H_b}{\partial a} - \frac{\partial H_a}{\partial b}\right)\frac{\partial}{\partial y}\left(\frac{1}{r}\right)d\tau - \iiint\left(\frac{\partial H_a}{\partial c} - \frac{\partial H_c}{\partial a}\right)\frac{\partial}{\partial z}\left(\frac{1}{r}\right)d\tau\right\}$$

$$= -\frac{1}{4\pi}\left\{\iiint\left(\frac{\partial H_b}{\partial a} - \frac{\partial H_a}{\partial b}\right)\frac{\partial}{\partial b}\left(\frac{1}{r}\right)d\tau - \iiint\left(\frac{\partial H_a}{\partial c} - \frac{\partial H_c}{\partial a}\right)\frac{\partial}{\partial c}\left(\frac{1}{r}\right)d\tau\right\}$$

= Surface Integrals

$$+ \frac{1}{4\pi}\left\{\iiint\frac{\partial}{\partial b}\left(\frac{\partial H_b}{\partial a} - \frac{\partial H_a}{\partial b}\right)\frac{d\tau}{r} - \iiint\frac{\partial}{\partial c}\left(\frac{\partial H_a}{\partial c} - \frac{\partial H_c}{\partial a}\right)\frac{d\tau}{r}\right\},$$

after an integration by parts. The surface can always be chosen so that the surface integrals vanish, and thus

$$\text{rot } U = \frac{1}{4\pi}\iiint\left\{\frac{\text{rot rot } H}{r}\right\}d\tau$$

$$= \frac{1}{4\pi}\iiint\frac{\text{grad div } H - \nabla^2 H}{r}\,d\tau.$$

Now the divergence of H is always zero so that

$$\text{rot } U = -\frac{1}{4\pi}\iiint\frac{\nabla^2 H}{r}\,d\tau \quad\ldots\ldots\ldots\ldots\ldots\ldots(28).$$

By comparing this with the equation for the potential

$$V = -\frac{1}{4\pi}\iiint\frac{\nabla^2 V}{r}\,d\tau,$$

we see that the right-hand side of (28) is equal to H. This proves the theorem, which is seen to be true for any vector whose divergence is always zero.

Having proved that a vector U such that $H = \text{rot } U$ can always be found, we substitute the new value of H in the equation

$$\text{rot } E = -\frac{1}{c}\frac{\partial H}{\partial t},$$

and obtain $$\text{rot}\left(E + \frac{1}{c}\frac{\partial U}{\partial t}\right) = 0.$$

This equation is satisfied if $E + \frac{1}{c}\frac{\partial U}{\partial t}$ is the gradient of some scalar quantity $-\phi$. So that we may put

$$E = -\frac{1}{c}\frac{\partial U}{\partial t} - \text{grad}\,\phi.$$

This is the same as equation (24). It will be observed, however, that U and ϕ are not completely determined by the considerations which have been brought forward. The only condition we have imposed on U except (24) is that it should satisfy the equation $H = \text{rot}\,U$. Also ϕ may be any scalar quantity. If U_0 and ϕ_0 are particular values of U and ϕ which satisfy the equations under consideration, they will also be satisfied by

$$U = U_0 - \text{grad}\,\psi \quad \text{and} \quad \phi = \phi_0 + \frac{1}{c}\frac{\partial \psi}{\partial t} \quad \text{.........(29)},$$

where ψ is some scalar function. We shall determine ψ by making it satisfy the condition

$$\text{div}\,U = -\frac{1}{c}\frac{\partial \phi}{\partial t} \quad \text{....................(30)}.$$

It is necessary to show that this condition can always be satisfied. Substituting the values (29) in equation (30) we get

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = \text{div}\,U_0 + \frac{1}{c}\frac{\partial \phi_0}{\partial t}.$$

There is always some value ψ which will satisfy this equation, so that (30) can always be satisfied.

We have

$$\rho = \text{div}\,E = -\frac{1}{c}\frac{\partial}{\partial t}(\text{div}\,U) - \text{div grad}\,\phi,$$

whence making use of (30)

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\rho.$$

We also have

$$\text{rot}\,H = \frac{1}{c}\left(\frac{\partial E}{\partial t} + \rho V\right).$$

Substituting from (23) and (24)

$$\text{rot rot}\,U = \text{grad div}\,U - \nabla^2 U$$
$$= \frac{1}{c}\left\{-\frac{1}{c}\frac{\partial^2 U}{\partial t^2} - \text{grad}\frac{\partial \phi}{\partial t} + \rho V\right\},$$

and
$$\operatorname{div} U = -\frac{1}{c}\frac{\partial \phi}{\partial t},$$

so that
$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = -\frac{1}{c}\rho V.$$

Thus (25) and (26) are the equations satisfied by the scalar and vector potentials respectively. Conversely if ϕ and U satisfy (25) and (26) the vectors E and H will be given by (23) and (24).

In the light of our discussion of equations (7) and (8) it is clear that the values of the potentials are

$$\phi = +\frac{1}{4\pi}\iiint \frac{\rho'}{r}\,d\tau$$

and
$$U = \frac{1}{4\pi c}\iiint \frac{(\rho V)'}{r}\,d\tau,$$

where the dashes denote that in carrying out the integrations the values of ρ and ρV, respectively, at the instant $t = -r/c$, previous to that for which the integrals are being evaluated, have to be substituted.

Electron at Rest and in Uniform Motion.

As an illustration of the results which we have just obtained we shall consider the case of a single electron. If the electron has always been at rest then V is always zero; so that the vector potential U vanishes. Moreover ρ' becomes identical with ρ for every point, since the position at any previous instant is the same as the instantaneous position. Thus the scalar potential is identical with the ordinary static potential, the electric intensity is identical with the usual value of electrostatics and the magnetic force vanishes. The solution in this case is identical with the results of the usual electrostatic theory.

Next consider an electron which is moving and has always moved with a uniform velocity w in a straight line parallel to the axis of z. Consider the values of the two potentials at any point P_1 at an instant t_1. They will not be determined by the instantaneous state of the electron, but by its state at some previous instant t_1'. t_1' will in fact be given by the equation

$$t_1' = t_1 - r_1'/c,$$

where r_1' may be called the retarded radius. It is of course a definite function of the instantaneous radius r_1, the velocity w and the velocity of light c. It is not, however, necessary for our present purpose that we should evaluate r_1' explicitly. Let us consider the potentials at a point P_2 such that P_1P_2 is parallel to the axis of z, and at a time t_2, such that $P_1P_2 = w(t_2 - t_1)$. Both the electron and the point P have moved forward in space a distance equal to $w(t_2 - t_1)$, and nothing else in the problem has changed. The potentials will have the same values at P_2 at the instant t_2 as they had at P_1 at the instant t_1; since they must be determined by the relative positions of P and the moving charge and cannot depend on their absolute positions in space. Thus the field due to a uniformly moving electron is carried along with it as though it were fixed to it by a rigid framework.

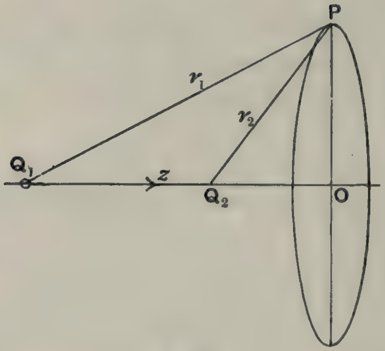


Fig. 28.

If Q_1O represents the direction of motion it is clear from symmetry that the potentials will have identical values at all points such as P which lie on a circle of given radius about a point O on the line of motion. Let $Q_1O = z$ and $Q_1P = r_1$, Q_1 being the instantaneous position of the electron at the time t for which the potentials at P are being calculated. The potentials will be functions of z, r_1 and t only. Moreover the resultant velocity lies along the axis of z, so that the x and y components of the vector potential vanish. We may therefore put

$$\phi = f(z, r_1, t),$$
$$U_x = U_y = 0,$$
$$U_z = F(z, r_1, t),$$

where f and F are functions which it is not necessary for us to evaluate more explicitly at present.

We have
$$E = -\frac{1}{c}\frac{\partial U}{\partial t} - \text{grad}\,\phi,$$

so that
$$E_x = -\frac{\partial\phi}{\partial x} = -\frac{x}{r_1}\frac{\partial f}{\partial r_1},$$

$$E_y = -\frac{\partial\phi}{\partial y} = -\frac{y}{r_1}\frac{\partial f}{\partial r_1},$$

$$E_z = -\frac{1}{c}\frac{\partial F}{\partial t} - \frac{\partial\phi}{\partial z}$$

$$= -\frac{1}{c}\frac{\partial F}{\partial t} - \frac{\partial f}{\partial z} - \frac{z}{r_1}\frac{\partial f}{\partial r_1}.$$

Thus the electric intensity may be regarded as made up of two components, the first parallel to the axis of z (the direction of motion of the electron) and equal to $-\frac{1}{c}\frac{\partial F}{\partial t} - \frac{\partial f}{\partial z}$ and the second directed along the instantaneous radius r_1 and equal to $-\frac{\partial f}{\partial r_1}$.

The components of the magnetic intensity are

$$H_x = \frac{\partial U_z}{\partial y} - \frac{\partial U_y}{\partial z} = \frac{y}{r_1}\frac{\partial F}{\partial r_1},$$

$$H_y = \frac{\partial U_x}{\partial z} - \frac{\partial U_z}{\partial x} = -\frac{x}{r_1}\frac{\partial F}{\partial r_1},$$

and
$$H_z = \frac{\partial U_y}{\partial x} - \frac{\partial U_x}{\partial y} = 0.$$

Thus the magnetic intensity is tangential to circles whose centres lie on the axis of motion and whose planes are perpendicular to that direction. The distribution of magnetic force is to this extent similar to that arising from a straight current lying on the axis of motion.

Accelerated Electron.

We shall next take an illustration in which the nature of the motion alters during the interval under consideration. Let the particle be at rest at the point Q_1 until the instant t_1; let it then be suddenly accelerated so that it acquires a finite velocity in an infinitesimal interval of time; then let it move with uniform velocity in a straight line until it reaches the point Q_2; at Q_2 it is stopped as suddenly as it was started at Q_1; let the stoppage be complete at the instant t_2. After the instant t_2 the particle remains for ever at rest in the position Q_2. Consider the field at P and let $PQ_1 = r_1$ and $PQ_2 = r_2$.

The field at P at any time t is not determined by the instantaneous position of Q but by its position at the time

$$t' = t - r'/c,$$

where r' is the retarded radius. Up to the instant when Q began to move r' was fixed and equal to r_1, so that provided t satisfies the inequality

$$t_1 > t - r_1/c,$$

or

$$t < t_1 + r_1/c,$$

the field at P will be that due to a static charge at Q_1. After the instant t_2 the radius r' becomes permanently equal to r_2 so that if

$$t_2 < t - r_2/c,$$

or

$$t > t_2 + r_2/c,$$

the field at P is that due to a static charge at Q_2. In the interval between $t = t_1 + r_1/c$ and $t = t_2 + r_2/c$ the field at P passes through three stages: (1) that due to a particle moving with a positive acceleration, (2) that due to a particle moving with uniform velocity, and (3) that due to a particle moving with a negative acceleration.

We shall see in the sequel that the field due to an accelerated electric charge possesses novel features of great interest.

CHAPTER X

THE ACTIVITY OF THE FORCES

WE shall next consider the rate at which work is done by the forces in the field, or, if we prefer this mode of expression, by the aether, on the charges in any given enclosed volume τ. The mechanical force on a unit charge is $E + \frac{1}{c}[VH]$, where V is the velocity of the charge relative to the instrument used in measuring H. The force acting on the electric charge in an element of volume $d\tau$ is therefore

$$\rho \left(E + \frac{1}{c}[VH] \right) d\tau,$$

and the rate at which work is being done by this force at any instant is equal to the scalar product of the resultant velocity of the element by the resultant force acting on it. The rate of working of the forces in the field on all the charges present in it is therefore

$$A = \iiint \rho \left(V \,.\, \overline{E + \frac{1}{c}[VH]} \right) d\tau.$$

Now the part of the electromotive intensity $\frac{1}{c}[VH]$ is always perpendicular to the plane containing V and H and is therefore always perpendicular to V. Thus $(V\,[VH])$ is always zero, so that the activity of the forces reduces to

$$A = \iiint (\rho V, E)\, d\tau.$$

But $\rho V = c \operatorname{rot} H - \frac{\partial E}{\partial t}$, so that

$$A = \iiint \left\{ c\,(\operatorname{rot} H E) - E\,\frac{\partial E}{\partial t} \right\} d\tau.$$

Replacing rot H by its Cartesian equivalents this becomes

$$\iiint\left\{c\left[E_x\left(\frac{\partial H_y}{\partial z}-\frac{\partial H_z}{\partial y}\right)+E_y\left(\frac{\partial H_z}{\partial x}-\frac{\partial H_x}{\partial z}\right)+E_z\left(\frac{\partial H_x}{\partial y}-\frac{\partial H_y}{\partial x}\right)\right]-E\frac{\partial E}{\partial t}\right\}d\tau.$$

If we collect together the terms containing $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ and $\frac{\partial}{\partial z}$ respectively and integrate each term by parts this becomes

$$c\iint dy\,dz\,(E_yH_z-E_zH_y)+dz\,dx\,(E_zH_x-E_xH_z)+dx\,dy\,(E_xH_y-E_yH_x)$$
$$-\iiint\left\{c\left[H_z\frac{\partial E_y}{\partial x}-H_y\frac{\partial E_z}{\partial x}+H_x\frac{\partial E_z}{\partial y}-H_z\frac{\partial E_x}{\partial y}+H_y\frac{\partial E_x}{\partial z}-H_x\frac{\partial E_y}{\partial z}\right]+E\frac{\partial E}{\partial t}\right\}d\tau.$$

If l, m, n are the direction cosines of an element of surface dS of the boundary of the volume τ we have $dy\,dz=l\,dS$, $dz\,dx=m\,dS$, $dx\,dy=n\,dS$, and $E_yH_z-E_zH_y$, $E_zH_x-E_xH_z$, $E_xH_y-E_yH_x$ are the x, y and z components respectively of the vector $[EH]$. Thus the integrated part is equal to

$$c\iint[EH]_n\,dS,$$

where $[EH]_n$ denotes the resolved part of the vector $[EH]$ along the normal to the element dS. The volume integral, after rearrangement, becomes

$$\iiint\left\{c\left[H_x\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)+H_y\left(\frac{\partial E_x}{\partial z}-\frac{\partial E_z}{\partial x}\right)+H_z\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right)\right]-E\frac{\partial E}{\partial t}\right\}d\tau,$$

and, since rot $E=\frac{1}{c}\frac{\partial H}{\partial t}$ and $H^2=H_x^2+H_y^2+H_z^2$, this may be written

$$-\iiint\frac{\partial}{\partial t}\{\tfrac{1}{2}H^2+\tfrac{1}{2}E^2\}\,d\tau.$$

Thus $$A=c\iint[EH]_n\,dS-\frac{\partial}{\partial t}\iiint(\tfrac{1}{2}H^2+\tfrac{1}{2}E^2)\,d\tau \quad ......(1).$$

Now $\frac{1}{2}H^2$ is the magnetic energy per unit volume of the aether and $\frac{1}{2}E^2$ is the electrostatic energy per unit volume. It is to be borne in mind that the E and H of this investigation are the universal values of the forces introduced in the last chapter and are not the same inside material media as either of the average values which we have previously defined as the intensities and inductions in such media. The mean value, over a region of the appropriate dimensions, of the present $\frac{1}{2}H^2$ is identical with the former $\frac{1}{2}\mu H^2$ or $\frac{1}{2}HB$ or $B^2/2\mu$, and the mean value of the present $\frac{1}{2}E^2$ is identical with the former $\frac{1}{2}\kappa E^2$ or $\frac{1}{2}ED$ or $D^2/2\kappa$. When E and H are the universal functions

$$-\frac{\partial}{\partial t}\iiint \tfrac{1}{2}(H^2+E^2)\,d\tau$$

represents the rate at which energy is lost by the aether, or space, within the limited region τ. Thus the whole of the work done by the forces of the field on the electric charges is not covered by the energy lost by the aether in the immediate neighbourhood. In general we have also to consider the quantity represented by the surface integral. Since the left-hand side of the activity equation is the rate at which work is done on the electric charges, and the volume integral represents the rate of loss of energy by the electromagnetic field in the enclosed volume, the surface integral which is equal to their difference must represent the rate at which energy flows into the region τ from outside. Any possible alternative to this conclusion would involve a denial of the principle of the Conservation of Energy.

Poynting's Theorem.

The occurrence of the surface integral $c\iint [EH]_n dS$ in the equation of activity of the forces was first remarked by Poynting[*], who gave to it a very definite physical interpretation. He pointed out that the behaviour of the field could be explained by the supposition that at every point there was a stream of energy equal per unit area to $c\,[E\,.\,H]$, the direction of the stream being coincident with that of this vector, and therefore normal to the plane containing E and H. The Poynting Flow of Energy thus vanishes when E and H are coincident in direction, and has a maximum value, other things being equal, when they are at right angles to one another. This interpretation is, of course, consistent with the equation of activity, and is, in fact, the most obvious interpretation of it. It is, however, not the only interpretation. For we may evidently add to $c[E.H]$ any vector R which satisfies the condition that the surface integral of its normal component over any closed surface vanishes: and the equation of activity will still be satisfied. We know from Gauss's Theorem that this condition will hold if we have $\operatorname{div} R = 0$ everywhere; so that there are an infinite number of vectors in addition to Poynting's which satisfy the equation of activity.

---

[*] *Phil. Trans.* A. vol. CLXXV. p. 343 (1884).

It is interesting to consider to what picture of the flow of energy we are led in typical instances on the supposition that it coincides with Poynting's vector. In the case of a straight wire carrying a current, for example, the electric force is parallel to the length of the wire, and the magnetic force is in circles about its axis. Thus the electric and magnetic forces are at right angles to one another, and the flow of energy is at right angles to both. That is to say, it flows perpendicularly into the wire from the insulating medium which surrounds it.

Probably the most convincing case of the flow of energy in accordance with Poynting's vector is that furnished by the propagation of electromagnetic waves. Consider a parallel beam of plane polarized light. We have seen that in such a beam the electric and magnetic vectors may be represented by

$$E = H = A\cos(pt - x).$$

It is important to notice that the two vectors are always in phase, and that they are equal in magnitude when expressed in the units used in this book. They are also at right angles to one another. Thus the resultant flow of energy is perpendicular to both E and H; that is to say, it is along the direction of propagation of the light. It is equal at any instant to

$$cA^2\cos^2(pt - x)$$

per unit area. Its average value over a single period, or over any very large interval of time, is

$$cA^2\frac{1}{\tau}\int_0^\tau \cos^2(pt - x)\,dt = \tfrac{1}{2}cA^2.$$

Now $\frac{1}{2}A^2$ is the mean energy present in unit volume of the wave, so that Poynting's Theorem represents this energy as flowing along with the velocity c. Since this is in accordance with the results of observation the application of Poynting's Theorem to questions relating to radiation evidently rests on very solid grounds.

Forces exerted on the Charges.

In the last section but one we have considered the rate of working or "activity" of the forces acting on an enclosed electrical system. We shall now consider the value of the resultant force acting on a similar system. On the electron theory of matter the results will be applicable to any material system since, on this theory, the force acting on a material system is the aggregate effect of the electric and magnetic forces which act on the electrons which constitute it. The force exerted on a unit charge is $E + \frac{1}{c}[VH]$, where V is the velocity of the charge relative to the system of instruments used to measure the forces. The reason for this particular specification of V will be clearer later (see Chaps. XIII and XIV). It will be observed that it is not inconsistent with the deduction from the magnetic properties of electric currents which led us to include the term $\frac{1}{c}[VH]$ in the expression for the force on a charged body (p. 114). So far as any evidence which has been considered up to the present is concerned, we might as well have taken V to be the velocity of the charge relative to the aether, which we might suppose to be absolutely fixed in space. When we come to consider the electrical and optical properties of bodies in very rapid motion, we shall see that the assumption that V refers to the velocity relative to the measuring system effects very important simplifications.

The charge present in the element of volume $d\tau$ being $\rho d\tau$, the force exerted on this element of volume will be

$$dF = \rho\left(E + \frac{1}{c}[VH]\right)d\tau,$$

and the force on the whole volume will be given by the vector equation

$$F=\iiint\rho\left(E+\frac{1}{c}[VH]\right)d\tau$$

$$=\iiint\left\{\operatorname{div}E\,.\,E+\frac{1}{c}[\rho V\,.\,H]\right\}d\tau,$$

since $\rho=\operatorname{div}E$. But

$$\operatorname{rot}H=\frac{1}{c}\left\{\frac{\partial E}{\partial t}+\rho V\right\};$$

hence
$$\rho V=c\operatorname{rot}H-\frac{\partial E}{\partial t},$$

and
$$F=\iiint\left\{\operatorname{div}E\,.\,E+\frac{1}{c}[c\operatorname{rot}H\,.\,H]-\frac{1}{c}\left[\frac{\partial E}{\partial t}\,.\,H\right]\right\}d\tau.$$

But
$$\left[\frac{\partial E}{\partial t}\,.\,H\right]=\frac{\partial}{\partial t}[EH]-\left[E\,.\,\frac{\partial H}{\partial t}\right]$$

$$=\frac{\partial}{\partial t}[EH]+c\,[E\,.\operatorname{rot}E],$$

$$[E\,.\operatorname{rot}E]=-[\operatorname{rot}E\,.\,E],$$

and
$$\operatorname{div}H=0;$$

hence
$$F=-\iiint\frac{1}{c}\frac{\partial}{\partial t}[EH]\,d\tau+\iiint\{\operatorname{div}E\,.\,E+[\operatorname{rot}E\,.\,E]\}\,d\tau$$

$$+\iiint\{\operatorname{div}H\,.\,H+[\operatorname{rot}H\,.\,H]\}\,d\tau\quad\ldots(2).$$

This is the total force on the volume τ. Consider the x component of F due to the third term of (2). Call it X_H. Then

$$X_H=\iiint\left\{\left(\frac{\partial H_x}{\partial x}+\frac{\partial H_y}{\partial y}+\frac{\partial H_z}{\partial z}\right)H_x+H_z\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)\right.$$
$$\left.-H_y\left(\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}\right)\right\}d\tau$$

$$=\iiint\left\{H_x\frac{\partial H_x}{\partial x}-H_y\frac{\partial H_y}{\partial x}-H_z\frac{\partial H_z}{\partial x}+H_x\frac{\partial H_y}{\partial y}+H_y\frac{\partial H_x}{\partial y}\right.$$
$$\left.+H_z\frac{\partial H_x}{\partial z}+H_x\frac{\partial H_z}{\partial z}\right\}d\tau$$

$$=\iiint\left\{\frac{\partial}{\partial x}\left[\tfrac{1}{2}(H_x^2-H_y^2-H_z^2)\right]+\frac{\partial}{\partial y}(H_xH_y)+\frac{\partial}{\partial z}(H_xH_z)\right\}d\tau$$

$$=+\iint dS\,\{l\,[\tfrac{1}{2}(H_x^2-H_y^2-H_z^2)]+m\,H_xH_y+n\,H_xH_z\},$$

where l, m, n are the direction cosines of the outward drawn normal to the element of area dS of the enclosing surface.

The third term of (2) will give rise to similar expressions for the components of force parallel to the y and z axes. If we treat the second term in (2) in the same way we shall get similar expressions except that the components of the magnetic intensity are replaced by the corresponding components of the electric intensity. If we put

$$p_{xx} = \tfrac{1}{2}\{E_x^2 - E_y^2 - E_z^2 + H_x^2 - H_y^2 - H_z^2\},\quad p_{yz} = p_{zy} = E_y E_z + H_y H_z,$$
$$p_{yy} = \tfrac{1}{2}\{E_y^2 - E_z^2 - E_x^2 + H_y^2 - H_z^2 - H_x^2\},\quad p_{zx} = p_{xz} = E_z E_x + H_z H_x,$$
$$p_{zz} = \tfrac{1}{2}\{E_z^2 - E_x^2 - E_y^2 + H_z^2 - H_x^2 - H_y^2\},\quad p_{xy} = p_{yx} = E_x E_y + H_x H_y,$$

the components of the last two terms of (2) may be written, selecting the x component as an example, in the form

$$\iint (l p_{xx} + m p_{yx} + n p_{zx})\, dS.$$

The last two terms therefore reduce to surface integrals over the boundary and are, in fact, identical with the forces due to the Maxwell stresses discussed in Chap. II, except for the addition of the magnetic terms which were not then being considered.

We now notice a very important difference between our present problem and the static case considered in Chap. II. The resultant of the forces acting on the volume from without, as calculated from the Maxwell stresses across the boundary, is no longer equal to the force tending to accelerate the charges enclosed by the boundary, the former being the greater by

$$\iiint \frac{1}{c}\frac{\partial}{\partial t}[EH]\, d\tau.$$

Thus, in the absence of electric charges, the resultant force due to the stresses over the boundary does not vanish unless the value of this volume integral is zero. The most natural interpretation would seem to be the following. In static cases, considering the action between the charges enclosed and the region external to the bounding surface, the resultant force on the external region, given by the Maxwell stresses, is equal and opposite to the resultant force on the charges enclosed. In general, however, the action and reaction between the charges and the external region are not equal and opposite; but part of the reaction to the force exerted on the external region falls on the aether enclosed by the surface, the amount of this reaction per unit volume being $\frac{1}{c}\frac{\partial}{\partial t}[EH]$. There are two other possibilities which are deserving of consideration. We may either deny the applicability of the Newtonian law of action and reaction to electrodynamic systems or we may deny the physical reality of the Maxwell stresses. Of these two alternatives the latter is to be preferred, at any rate from the standpoint of the electron theory of matter. For if we are to regard material systems as electrodynamic systems (which is the electron theory in a nutshell) it is essential that the Newtonian laws should be true for them to the extent that they are true for material systems. That is to say, they must be exceedingly close approximations to the truth in the case of systems the parts of which have velocities which are small compared with that of light.

It happens that we can retain the law of action and reaction and also the physical existence of the Maxwell stresses if we interpret the vector $[EH]/c$ as momentum per unit volume of the medium. In that case equation (2) shows that the stresses acting across the boundary of the region are equal to the force tending to move the charges within the region plus the rate at which momentum is communicated to the enclosed medium. The idea of electromagnetic momentum was first suggested by J. J. Thomson*. The momentum per unit volume is very closely related to Poynting's vector, $c[EH]$, being, in fact, equal to the latter divided by the square of the velocity of light.

We shall see that this idea of electromagnetic momentum is extremely useful in enabling us to make calculations about the mechanical effects of light and of moving electrons. At the same time it suffers from a serious disadvantage in so far as we have no satisfactory conception of any mode of motion of the aether to which it corresponds. Thomson has suggested† that it represents the inertia of tubes of electric force. These are supposed by him to have a definite concrete physical existence. Each tube is supposed to be anchored at one end to the electron to which it

* *Recent Researches in Electricity and Magnetism*, p. 13 (1893).

† *Loc. cit.*

belongs, whilst the other end extends to an indefinite distance. In a recent paper Thomson has examined the consequences which would follow if the number of such tubes attached to each electron were quite small*. It will be observed that this theory attaches the inertia in reality to the electrons rather than to the medium in which they move.

The most important application of the idea of electromagnetic momentum is to the dynamics of a moving electric charge. We shall, however, defer that question until the next chapter and occupy ourselves for the present with the question of the pressure exerted by light and other electromagnetic waves. This subject affords excellent illustrations of the application of the ideas both of electromagnetic momentum and of the aethereal stresses.

The Pressure of Radiation.

We shall consider the pressure of radiation first of all from the point of view of the Maxwell stresses. Let us apply equation (2) to any closed surface containing matter or electrons and consider the average value of the quantities occurring in the equation, taken over a considerable interval of time T. The average value of the left-hand side will be the average force exerted on the matter or electrons. There are important cases in which the average effect of the Maxwell stresses can be very easily calculated and in which the volume integral vanishes. Considering the last-named term first, its average value over an interval of time T is

$$\frac{1}{T}\int_0^T dt \iiint -\frac{1}{c}\frac{\partial}{\partial t}[EH]\,d\tau$$

$$= -\frac{1}{cT}\left|\iiint [EH]\,d\tau\right|_0^T.$$

This expression is equal to zero when either of the following conditions holds:—

(1) The volume integral has the same value at both the time limits. This will be the case when the electromagnetic actions are periodic and T is an integral multiple of the periodic time.

(2) The volume integral $\iiint [EH]\,d\tau$ is finite throughout

---

* *Phil. Mag.* vol. XIX. p. 301 (1909).

the time considered and T is an interval of time so great that it may be regarded as infinitely large.

It is evident that the electromagnetic momentum is relatively unimportant except when rapid changes are taking place in the state of the electromagnetic field.

Let us now apply our results to the case of a plane polarized wave of monochromatic light incident normally on a surface which absorbs it completely. Let $ABCD$ represent the absorbing substance, the light being incident on it in the direction of the arrow which is parallel to the z axis. Let the electric intensity in the light wave lie along the axis of x and be equal to $E_0 \cos pt$. Then the components of the electric and magnetic intensities are

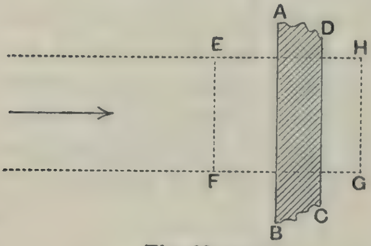


Fig. 29.

$$E_x = E = E_0 \cos pt,\ \ E_y = 0,\ E_z = 0,$$

$$H_x = 0,\ \ H_y = E_x = E_0 \cos pt = H = E,\ \ H_z = 0.$$

Applying equation (2) to the cylinder whose cross section by the plane of the paper is $EFGH$ and integrating over a complete period, we see that the part coming from the volume integral on the right-hand side vanishes. The left-hand side is equal to the average force exerted by the light on the matter in $ABCD$ by which it is absorbed. The components of the Maxwell stresses are :—

$$p_{xx} = \tfrac{1}{2}\{E_x^2 - E_y^2 - E_z^2 + H_x^2 - H_y^2 - H_z^2\} = 0,$$

$$p_{yy} = \tfrac{1}{2}\{E_y^2 - E_z^2 - E_x^2 + H_y^2 - H_z^2 - H_x^2\} = 0,$$

$$p_{zz} = \tfrac{1}{2}\{E_z^2 - E_x^2 - E_y^2 + H_z^2 - H_x^2 - H_y^2\} = -E^2,$$

$$p_{xy} = p_{yx} = E_x E_y + H_x H_y = 0.$$

Similarly $$p_{yz} = p_{xz} = 0.$$

Thus there is no stress on the cylindrical surface of which EH and FG are sections, since this surface is everywhere parallel to the axis of z. The stress vanishes over GH since there is no light there. The only part of the surface over which the Maxwell stresses are effective is the end EF, and they are here equivalent to a pressure E^2 per unit area. Thus the force exerted by the light on the opaque surface is a pressure which is equal per unit area to

$$\frac{1}{T}\int E^2\,dt = \frac{1}{T}\int E_0^2 \cos^2 pt\,dt = \frac{E_0^2}{2}.$$

The energy per unit volume of the beam of light is

$$\tfrac{1}{2}\{E^2 + H^2\} = E^2,$$

so that the intensity of the light, which is equal to the average amount of energy transported across unit area in unit time, is

$$\frac{1}{T}\int_0^T cE^2\,dt = \frac{cE_0^2}{2}.$$

Thus in the case of light incident normally on a perfectly absorbing surface the radiation pressure is equal to the intensity of the light divided by its velocity of propagation.

It is clear that in the case of a perfectly reflecting surface the value of the radiation pressure will be doubled since the *intensity* of the light at EF will be twice as great as with a perfectly absorbing surface, when the incident intensity is the same. Since these results are independent of the plane of polarization of the light they will also be true, at normal incidence, when the light is unpolarized. It follows, on similar grounds, that they are also true for light of mixed frequencies.

Isotropic Radiation.

Next consider a perfectly reflecting enclosure filled with isotropic radiation. By isotropic radiation we mean radiation which is being propagated in all directions in such a manner that the probability of the direction of propagation of any ray, selected at random, being found within a given solid angle is proportional to that solid angle. The thermal radiation which would fill the enclosure in the final state of equilibrium which ensues when there are material bodies within it is of this character (see Chap. XV). Consider an infinitesimal area dS of the reflecting enclosure and apply equation (2) to a cylinder, similarly situated to that in Fig. 29, but whose height is infinitesimal compared with the dimensions of its ends. As before, let dS be perpendicular to the axis of z. We may now neglect the tractions on the sides of the cylinder on account of the smallness of their dimensions. The components of the force arising from the Maxwell stresses therefore vanish everywhere except on the front face of the cylinder, and are there equal respectively to

$$p_{zz}dS = \tfrac{1}{2}\{E_z^2 - E_x^2 - E_y^2 + H_z^2 - H_x^2 - H_y^2\}\,dS,$$
$$p_{xz}dS = \{E_xE_z + H_xH_z\}\,dS,$$
$$p_{yz}dS = \{E_yE_z + H_yH_z\}\,dS.$$

Since by hypothesis the trains of waves which constitute the radiation are as likely to be travelling in any one direction as in any other, the electric and magnetic intensities are as likely to lie in any one direction as in any other and, incidentally, are as likely to be positive as negative. We therefore have the following relations among the average values of these quantities, taken over a long interval of time, for the radiation which crosses the end of the cylinder :—

$$\overline{E_x^2} = \overline{E_y^2} = \overline{E_z^2} = \tfrac{1}{3}\,\overline{E_x^2 + E_y^2 + E_z^2} = \tfrac{1}{3}\,\overline{E^2},$$
$$\overline{H_x^2} = \overline{H_y^2} = \overline{H_z^2} = \tfrac{1}{3}\,\overline{H_x^2 + H_y^2 + H_z^2} = \tfrac{1}{3}\,\overline{H^2},$$
$$\overline{E_xE_z} = \overline{E_yE_z} = \overline{H_xH_z} = \overline{H_yH_z} = 0.$$

Hence the mean values of the tractions are

$$\overline{p_{xz}}dS = \overline{p_{yz}}dS = 0,$$

and $$\overline{p_{zz}}dS = -\tfrac{1}{6}\{\overline{E^2} + \overline{H^2}\}\,dS = -\tfrac{1}{3}\,\overline{E^2}\,dS.$$

Here $\overline{E^2}\,(=\overline{H^2})$ is the mean square of the electric intensity in the radiation in the enclosure. Thus in the case of isotropic radiation the pressure is equal to one-third of the energy in unit volume of the radiation. This result has an important application to the Thermodynamics of radiation.

Radiation Pressure and Momentum.

Let us now consider the pressure of radiation from the point of view of the electromagnetic momentum. Confining our attention first of all to a plane wave of plane-polarized monochromatic radiation, we observe that this consists of an alternating electric and magnetic field which travels forward with the uniform velocity of light. As we have seen, the electric and magnetic intensities are always at right angles to each other and to the direction of propagation of the beam. In our units they are also equal in magnitude at any particular point at any instant, and for the same point in space they are harmonic functions of the time. If E and H are the values of the two intensities at any instant it follows that the electromagnetic momentum per unit volume is

$$\frac{1}{c}[EH] = \frac{1}{c} EH = \frac{1}{c} E^2 = \frac{1}{c} E_0^2 \cos^2 pt.$$

Thus a plane wave (whether polarized or not) of light of frequency $p/2\pi$ may be regarded as a stream of momentum which moves through the medium with the velocity of light. At a fixed point in the medium the momentum per unit volume is $\frac{1}{c} E_0^2 \cos^2 pt$. This momentum is always in the direction of propagation of the light, since E and H always have the same sign.

Now suppose that a train of plane waves, such as we have been considering, falls normally on a perfectly absorbing surface. All of the radiation disappears so that there is a constant stream of momentum flowing into the absorbing surface. But interchange of momentum implies the existence of a stress. The radiation will therefore exert a pressure on the surface which will be equal to the rate of change of momentum per unit area per unit time. The average pressure is therefore

$$\frac{1}{T}\int_0^T c \times \frac{1}{c} E_0^2 \cos^2 pt\, dt = \tfrac{1}{2} E_0^2,$$

in agreement with the value found previously. In the case where the surface is perfectly reflecting the forward momentum will not merely be destroyed but an equal and opposite momentum will be given to the reflected wave. Thus in this case there is twice as great a rate of change of momentum; so that the pressure will be twice as great as with a perfectly absorbing surface, for the same intensity in the incident light.

In the case when a plane wave is incident in a direction making an angle θ with the normal to a perfectly reflecting surface, it is convenient to resolve the momentum in the incident and reflected waves into two parts, one normal to the reflecting surface, and the other parallel to it. The momentum parallel to the normal which falls on unit area in unit time is $(1/c) S \cos^2 \theta$, where S is Poynting's vector $c\,[EH]$. The momentum in the same direction, which leaves unit area in unit time owing to the reflected wave, is

$-(1/c) S \cos^2\theta$. The factor $\cos\theta$ enters twice because, in the first place, the resultant momentum is along the direction of propagation of the wave and we are dealing with the component at an angle θ with this direction and, in the second place, on account of the radiation falling obliquely on the surface, unit area of the wave front will be spread over an area $1/\cos\theta$ of the surface. The rate of change of momentum normal to the surface is evidently $(2/c) S \cos^2\theta$ per unit area. Thus there is a normal pressure whose average value is $(2/c) \cos^2\theta\, \bar{S} = E_0^2 \cos^2\theta$.

The momentum parallel to the surface lies in the plane of incidence. The amount of it which is incident on unit area in unit time is clearly $(1/c) S \sin\theta \cos\theta$. The amount leaving unit area in unit time is also equal to this, since the direction of this part of the momentum is unchanged on reflexion. The rate of change of the tangential component of the momentum is therefore zero, so that there is no tangential stress, even when the radiation is incident obliquely, at the surface of a perfect reflector. It is evident that this conclusion is no longer true when part of the radiation is absorbed at the reflecting surface. There will then be a tangential stress which is proportional to the difference between the intensities of the incident and the reflected waves.

When light is incident at a transparent surface, we have to deal with a refracted as well as a reflected beam. Let θ_1 denote the angles of incidence and reflexion and θ_2 the angle of refraction. In this case we are no longer dealing with the free aether, so that we have to use the expression for the momentum per unit volume appropriate to a material substance. By making the changes in the argument on p. 205 which are necessary when dealing with media whose dielectric constant and magnetic permeability are not equal to unity, we see that the general expression for the momentum per unit volume is $[DB]/c$, where D is the electric and B the magnetic induction. Denoting the values of this quantity for the incident, refracted and reflected waves by G_1, G_2 and G_3 respectively, it is clear that the instantaneous value of the normal pressure is

$$V_1 (G_1 + G_3) \cos^2\theta_1 - V_2 G_2 \cos^2\theta_2,$$

where V_1 and V_2 are the velocities of radiation in the two media. The instantaneous value of the tangential stress is

$$V_1 (G_1 - G_3) \sin\theta_1 \cos\theta_1 - V_2 G_2 \sin\theta_2 \cos\theta_2.$$

Here $\sin\theta_1/\sin\theta_2 = V_1/V_2$ and the values of G_1, G_2 and G_3 are proportional to the respective instantaneous intensities of the beam.

The ratios of the average values of these can therefore be obtained from the expressions for the reflected and refracted intensities found in Chap. VII, and will depend on the plane of polarization of the light.

It remains to add that the pressure due to electromagnetic radiation was predicted by Maxwell[*], as a consequence of his theory of stresses in the medium, in 1873. It was first demonstrated by experiment by Lebedew[†] in 1899, and later independently by Nichols and Hull[‡] in 1901. Several of the more complex cases of the effect of light pressure have recently been examined by Poynting[§].

Isolated System.

An interesting application of the theorem expressed in equation (2) arises in the case of an isolated system. Take any surface surrounding the system, the dimensions of the enclosing surface being so great that the field at any point of it may be considered negligible. Then the Maxwell stresses vanish over the boundary and we have

$$\iiint \rho \left\{ E + \frac{1}{c}[VH] \right\} d\tau = -\frac{\partial}{\partial t}\iiint G\,d\tau.$$

But the left-hand side is the force acting on the charged bodies in the system and is, therefore, equal to the rate of increase of (supposedly) material momentum of this part of the system. Calling this momentum M we therefore have

$$\frac{\partial M}{\partial t} = -\frac{\partial}{\partial t}\iiint G\,d\tau,$$

or the momentum gained by the material part of the charged system is equal to the momentum lost by the electric field. We shall see in the next chapter that if matter is made up solely of electrons,

---

[*] *Treatise on Electricity and Magnetism*, § 792 (1873).

[†] *Arch. des Sciences Phys. et Nat.* (4), vol. VIII. p. 184 (1899); *Ann. der Phys.* vol. VI. p. 433 (1903).

[‡] *Phys. Rev.* vol. XIII. p. 293 (1901).

[§] *Phil. Mag.* vol. IX. pp. 169, 475 (1905).

M consists of a distribution of momentum in the field of the same type in reality as G. From this point of view the effect of the forces on the parts of the system may be regarded as giving rise to a rearrangement of the electromagnetic momentum in the field, the total amount remaining constant throughout the changes which take place. This result is only true provided the motions may be regarded as quasi-stationary (see p. 262); otherwise momentum will be lost by radiation over the boundary.

At present we can only be certain that matter is made up in part of electrons. If this should turn out to be ultimately true we should have to say that during quasi-stationary dynamical actions the electromagnetic momentum of the field is converted partly into the electromagnetic momentum of the individual electrons and partly into material momentum, the total momentum being unaltered.

CHAPTER XI

CHARGED SYSTEM IN UNIFORM MOTION

We have seen that a charged system in uniform motion carries its field along with it as though it were rigidly attached to it and that the potentials and forces are given by the equations

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -\rho,$$

$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = -\rho V,$$

$$E = -\frac{1}{c}\frac{\partial U}{\partial t} - \operatorname{grad}\phi,$$

$$H = \operatorname{rot} U.$$

We shall suppose the charged system to be moving along the axis of z with the uniform velocity w, and that it is symmetrical about the axis of motion. The equations above refer to axes fixed relatively to the observer, past whom the system is moving uniformly with velocity $V = w$. If we consider axes which move along with the system we can take advantage of the fact that the field is invariable relative to points measured along these axes. If $\frac{D}{Dt}$ denotes the rate of change of any quantity with respect to time at a point x, y, z fixed with respect to the moving axes, then

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z},$$

where $\partial/\partial t$ denotes differentiation at a point fixed with respect to the fixed axes, and u, v, w are the velocity components of the moving axes. In the present case there is no change with time for a point moving with the moving axes, so that $\frac{D}{Dt}=0$. We also have $u=v=0$. So that

$$\frac{\partial}{\partial t}=-w\frac{\partial}{\partial z},$$

and
$$\frac{\partial^2}{\partial t^2}=w^2\frac{\partial^2}{\partial z^2}.$$

Thus the fact that the field is carried along by the moving system as though it were rigidly attached to it enables us to eliminate the time from the equations for ϕ and U. On substituting for $\frac{\partial^2}{\partial t^2}$, the equations for ϕ and U become

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+(1-\beta^2)\frac{\partial^2\phi}{\partial z^2}=-\rho,$$

and
$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+(1-\beta^2)\frac{\partial^2 U}{\partial z^2}=-\beta\rho,$$

where $\beta=w/c$. By changing the variables to x_1, y_1, z_1, where

$$x_1=x,\quad y_1=y,\quad z_1=z/\sqrt{1-\beta^2},$$

these equations may be written

$$\nabla_1^2\phi\equiv\frac{\partial^2\phi}{\partial x_1^2}+\frac{\partial^2\phi}{\partial y_1^2}+\frac{\partial^2\phi}{\partial z_1^2}=-\rho \quad \dots\dots\dots\dots\dots(1),$$

$$\nabla_1^2 U\equiv\frac{\partial^2 U}{\partial x_1^2}+\frac{\partial^2 U}{\partial y_1^2}+\frac{\partial^2 U}{\partial z_1^2}=-\beta\rho \quad \dots\dots\dots\dots(2).$$

These equations are of the same form as Poisson's equation. Thus the scalar and vector potentials due to a moving charged system may be obtained from the scalar potential for a slightly different system at rest. The transformed system is obtained from the moving system by stretching all the lengths in the moving system parallel to the direction of motion in the ratio 1 to $\sqrt{1-\beta^2}$. This type of transformation was first obtained by J. J. Thomson* by a different method and also, independently, by Heaviside†.

It will be noticed that the scalar and vector potentials are proportional to one another, differing only by the constant factor β. It will therefore be necessary for us to calculate only one of them.

---

* *Phil. Mag.* April 1881, also *Recent Researches in Electricity and Magnetism*, p. 19.

† *Phil. Mag.* April 1889.

Let us now denote our moving system by S and suppose that corresponding to S there is a fixed system S_1 which is obtained by enlarging the dimensions of the moving system S in the ratio $1 : \sqrt{1-\beta^2}$ in the direction of motion. Every point x, y, z in the moving system will correspond to a point x_1, y_1, z_1 in the fixed system, where the relation between x, y, z and x_1, y_1, z_1 is that already given. Also, corresponding elements of volume will contain equal charges, since straining the dimensions does not alter the quantity of electricity present. Thus, if ρ is the volume density in the moving system and ρ_1 the corresponding quantity in the corresponding element of volume in the fixed system, we have the equation

$$\rho\, dx\, dy\, dz = \rho_1\, dx_1\, dy_1\, dz_1.$$

But $dx = dx_1$, $dy = dy_1$, and $dz = dz_1 \sqrt{1-\beta^2}$;

hence $$dx\, dy\, dz = dx_1\, dy_1\, dz_1 \sqrt{1-\beta^2},$$

and $$\rho_1 = \rho \sqrt{1-\beta^2} \quad \text{..........................(3).}$$

The potential ϕ_1 of the distribution in the fixed system must satisfy Poisson's equation

$$\frac{\partial^2 \phi_1}{\partial x_1^2} + \frac{\partial^2 \phi_1}{\partial y_1^2} + \frac{\partial^2 \phi_1}{\partial z_1^2} = -\rho_1 = -\sqrt{1-\beta^2}\,.\,\rho \quad \text{.........(4).}$$

But the scalar potential ϕ which we are seeking, satisfies the equation

$$\frac{\partial^2 \phi}{\partial x_1^2} + \frac{\partial^2 \phi}{\partial y_1^2} + \frac{\partial^2 \phi}{\partial z_1^2} = -\rho,$$

whence $$\phi_1 = \sqrt{1-\beta^2}\,.\,\phi \quad \text{.........................(5).}$$

Thus the scalar potential in the moving system is equal to $1/\sqrt{1-\beta^2}$ times the electrostatic potential in the corresponding fixed system.

It is desirable to emphasize at this point that the fixed system which we have imagined is simply a mathematical device to facilitate the calculations. We are not supposing that the moving system is transformed physically in any way into the corresponding fixed system. What we have proved is that the scalar and vector potentials of the moving system are related in a certain way to the electrostatic potential of the corresponding fixed system, and it is comparatively easy to calculate the value of the latter. I have mentioned this because I find that students who have a little knowledge of the principle of relativity are apt to become confused as to the point at issue.

We are now in a position to express the energy and the electromagnetic momentum in the moving system S in terms of the potential ϕ_1 and the coordinates x_1, y_1, z_1 of the corresponding fixed system S_1. The electric energy per unit volume is

$$\tfrac{1}{2}\{E_x^2+E_y^2+E_z^2\}.$$

We notice that

$$\frac{\partial\phi}{\partial x}=\frac{\partial\phi}{\partial x_1}=\frac{1}{\sqrt{1-\beta^2}}\frac{\partial\phi_1}{\partial x_1},\qquad \frac{\partial\phi}{\partial y}=\frac{1}{\sqrt{1-\beta^2}}\frac{\partial\phi_1}{\partial y_1},$$

$$\frac{\partial\phi}{\partial z}=\frac{1}{\sqrt{1-\beta^2}}\frac{\partial\phi}{\partial z_1}=\frac{1}{1-\beta^2}\frac{\partial\phi_1}{\partial z_1}.$$

We have seen that the resultant vector potential is parallel to the direction of motion. We thus have

$$U_x=U_y=0 \text{ and } U_z=\beta\phi.$$

Hence $E_x=-\dfrac{\partial\phi}{\partial x}=-\dfrac{1}{\sqrt{1-\beta^2}}\dfrac{\partial\phi_1}{\partial x_1}$, $\qquad E_y=-\dfrac{1}{\sqrt{1-\beta^2}}\dfrac{\partial\phi_1}{\partial y_1}$,

$$E_z=-\frac{1}{c}\frac{\partial}{\partial t}(U_z)-\frac{\partial\phi}{\partial z}=\frac{w}{c}\frac{\partial}{\partial z}(\beta\phi)-\frac{\partial\phi}{\partial z}$$

$$=-(1-\beta^2)\frac{\partial\phi}{\partial z}=-\frac{\partial\phi_1}{\partial z_1}.$$

The electric energy of the moving system S is

$$W=\frac{1}{2}\iiint\{E_x^2+E_y^2+E_z^2\}\,dx\,dy\,dz$$

$$=\frac{1}{2}\iiint\left\{\frac{1}{1-\beta^2}\left[\left(\frac{\partial\phi_1}{\partial x_1}\right)^2+\left(\frac{\partial\phi_1}{\partial y_1}\right)^2\right]+\left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\}\sqrt{1-\beta^2}\,dx_1\,dy_1\,dz_1$$

$$=\frac{1}{2}\iiint\left\{\frac{1}{\sqrt{1-\beta^2}}\left[\left(\frac{\partial\phi_1}{\partial x_1}\right)^2+\left(\frac{\partial\phi_1}{\partial y_1}\right)^2\right]+\sqrt{1-\beta^2}\left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\}dx_1\,dy_1\,dz_1 \qquad \text{......(6).}$$

The components of the magnetic intensity are

$$H_x = \frac{\partial U_z}{\partial y} - \frac{\partial U_y}{\partial z} = \beta \frac{\partial \phi}{\partial y} = \frac{\beta}{\sqrt{1-\beta^2}} \frac{\partial \phi_1}{\partial y_1},$$

$$H_y = \frac{\partial U_x}{\partial z} - \frac{\partial U_z}{\partial x} = -\beta \frac{\partial \phi}{\partial x} = -\frac{\beta}{\sqrt{1-\beta^2}} \frac{\partial \phi_1}{\partial x_1},$$

$$H_z = \frac{\partial U_y}{\partial x} - \frac{\partial U_x}{\partial y} = 0.$$

The magnetic energy of S is therefore

$$\begin{aligned} T &= \frac{1}{2}\iiint \{H_x^2 + H_y^2 + H_z^2\}\, dx\,dy\,dz \\ &= \frac{1}{2}\iiint \frac{\beta^2}{1-\beta^2}\left\{\left(\frac{\partial \phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial \phi_1}{\partial x_1}\right)^2\right\}\sqrt{1-\beta^2}\, dx_1\,dy_1\,dz_1 \\ &= \frac{1}{2}\frac{\beta^2}{\sqrt{1-\beta^2}}\iiint\left\{\left(\frac{\partial \phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \phi_1}{\partial y_1}\right)^2\right\} dx_1\,dy_1\,dz_1 \quad \text{............(7).}\end{aligned}$$

The components of Poynting's vector $S = c\,[EH]$ are

$$\begin{aligned} S_x &= c\,(E_y H_z - E_z H_y) \\ &= -c\,\frac{\beta}{\sqrt{1-\beta^2}} \frac{\partial \phi_1}{\partial x_1}\frac{\partial \phi_1}{\partial z_1} \quad \text{........................(8),}\end{aligned}$$

$$\begin{aligned} S_y &= c\,(E_z H_x - E_x H_z) \\ &= -c\,\frac{\beta}{\sqrt{1-\beta^2}} \frac{\partial \phi_1}{\partial y_1}\frac{\partial \phi_1}{\partial z_1} \quad \text{........................(9),}\end{aligned}$$

$$\begin{aligned} S_z &= c\,(E_x H_y - E_y H_x) \\ &= c\,\frac{\beta}{1-\beta^2}\left\{\left(\frac{\partial \phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \phi_1}{\partial y_1}\right)^2\right\} \quad \text{..............(10).}\end{aligned}$$

We have seen that the momentum per unit volume is $\frac{1}{c^2}S$, so that the components of the total electromagnetic momentum of the system S are

$$\begin{aligned} G_x &= -\frac{\beta}{c\sqrt{1-\beta^2}}\iiint \frac{\partial \phi_1}{\partial x_1}\frac{\partial \phi_1}{\partial z_1}\, dx\,dy\,dz \\ &= -\frac{\beta}{c}\iiint \frac{\partial \phi_1}{\partial x_1}\frac{\partial \phi_1}{\partial z_1}\, dx_1\,dy_1\,dz_1 \quad \text{......................(11),}\end{aligned}$$

$$G_y = -\frac{\beta}{c}\iiint \frac{\partial \phi_1}{\partial y_1}\frac{\partial \phi_1}{\partial z_1}\, dx_1\,dy_1\,dz_1 \quad \text{......................(12),}$$

$$G_z = +\frac{\beta}{c\sqrt{1-\beta^2}}\iiint\left\{\left(\frac{\partial \phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \phi_1}{\partial y_1}\right)^2\right\} dx_1\,dy_1\,dz_1 \quad \text{...(13).}$$

The results we have obtained are true for any electrical system in uniform motion. We shall apply them to two simple cases which we shall select partly on the ground of their simplicity, but also because we may reasonably expect that they will give us an insight into the behaviour of the electron. The case that we shall consider first is that of a rigid thin spherical shell of uniform electrification.

The Rigid Electron.

We shall now suppose that the moving system consists of a spherical shell of radius R which is uniformly electrified. The thickness of the shell will be supposed to be negligible compared with its radius. It follows from the results that have already been established, that the energy, momentum, etc. of the field due to such an electrified shell when in motion may be obtained if we can calculate the static potential ϕ_1 of the corresponding system which arises when the actual space of the problem is strained so that all lengths parallel to the direction of motion are increased in the ratio $1 : \sqrt{1-\beta^2}$, whilst lengths at right angles to this remain unchanged. The moving sphere will evidently strain into an ellipsoid of revolution in the corresponding fixed system. The axis of revolution of the ellipsoid coincides with the direction of motion of the sphere. The major axis of any of the principal elliptic sections is equal to $R/\sqrt{1-\beta^2}$, whilst the minor axis is equal to the radius R of the moving sphere.

It is to be observed that these results will only be true provided the electrified sphere is rigid. Certain experiments, which will be discussed later, have led physicists to suspect that the lengths of bodies depend on their velocities relative to that of the observer engaged in measuring them. If this kind of change affects the electron itself, as well as the aggregate of electrons which we suppose constitutes the material substance, the figure in the fixed system which corresponds to the moving sphere will no longer be the ellipsoid which we have described. This follows because the moving system which we suppose to be spherical when at rest becomes distorted, and is no longer spherical when in motion. We shall see later that if the shape of the charged sphere does change, and if the changes are such as would naturally be suggested by the results of the experiments, the calculations become much simpler than those which we are now carrying out for the rigid spherical shell.

In order to make our notation agree better with that usual in the geometry of ellipsoids we shall suppose the motion to be parallel to the axis of x instead of that of z. The appropriate changes in the formulae of the last section may easily be made on inspection. The equation of the ellipsoid into which the moving sphere distorts is

$$\frac{x_1^2}{a^2} + \frac{y_1^2 + z_1^2}{b^2} = 1 \quad \text{.....................(14)},$$

where $a = R/\sqrt{1 - \beta^2}$ and $b = R$ are the semi-axes major and minor respectively. The equation of the family of ellipsoids which are confocal with (14) may be written

$$\frac{x_1^2}{p^2 + \lambda} + \frac{y_1^2 + z_1^2}{\lambda} = 1 \text{.....................(15)},$$

if $p^2 = a^2 - b^2 =$ the square of half the distance between the foci. The spheres which bound the rigid electrified shell will transform into two infinitely near, similar and similarly placed ellipsoids, one of which is given by the equation (14). The space between the similar ellipsoids is filled with a distribution of electrification of uniform density. The potential due to such a distribution is constant within the region bounded by the ellipsoidal shell, the distribution being equivalent to that on an ellipsoidal conductor maintained at a constant potential. Outside the shell the equipotential surfaces are the confocal ellipsoids given by equation (15)*. The difference of potential between two confocal ellipsoids whose equations are given by

$$\frac{x^2}{\lambda_3^2} + \frac{y^2 + z^2}{\lambda_3^2 - c^2} = 1$$

is

$$V_1 - V_2 = e\,\frac{\gamma_1 - \gamma_2}{c},\dagger$$

where

$$\gamma = \int_c^{\lambda_3} \frac{c\,d\lambda_3}{\lambda_3^2 - c^2},$$

and e is the charge on the inner ellipsoid. Applying this result to the ellipsoids (14) and (15) we have $\lambda_3 = \sqrt{p^2+\lambda}$ corresponding to (15), $\lambda_3' = a$ corresponding to (14), $c = p = \sqrt{a^2-b^2}$ and

---

* Cf. Webster, *Electricity and Magnetism*, Chap. v.

† Maxwell, *Electricity and Magnetism*, 2nd ed. vol. I. p. 237, equation (28).

$$\gamma_1 - \gamma_2 = \frac{1}{2}\left\{\log\frac{\sqrt{p^2+\lambda}-p}{\sqrt{p^2+\lambda}+p} - \log\frac{a-p}{a+p}\right\}.$$

By removing the outer ellipsoid to an infinite distance and introducing a factor $1/4\pi$ on account of the difference between our units of electric charge and those used by Maxwell, we find that the potential at any point in any ellipsoid confocal with (14) and given by equation (15) is

$$\phi_1 = \frac{e}{8\pi p}\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p} \quad\text{..................(16).}$$

Since the potential is the same at every point of the ellipsoidal shell, the total energy of its electrostatic field is

$$\tfrac{1}{2}e\phi_0 = \frac{e^2}{16\pi\sqrt{a^2-b^2}}\log\frac{a+\sqrt{a^2-b^2}}{a-\sqrt{a^2-b^2}}.$$

This must also be equal to

$$\frac{1}{2}\iiint\left\{\left(\frac{\partial\phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\} dx_1\,dy_1\,dz_1.$$

We have seen that the energy and momentum of the moving system depend upon the integrals

$$J_1 = \iiint\left(\frac{\partial\phi_1}{\partial x_1}\right)^2 dx_1\,dy_1\,dz_1 \quad\text{......................(17),}$$

and

$$J_2 = \iiint\left\{\left(\frac{\partial\phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\} dx_1\,dy_1\,dz_1 \quad\text{.........(18).}$$

We therefore have one simple relation between them, namely

$$\tfrac{1}{2}(J_1+J_2) = \tfrac{1}{2}e\phi_0 = \frac{e^2}{16\pi\sqrt{a^2-b^2}}\log\frac{a+\sqrt{a^2-b^2}}{a-\sqrt{a^2-b^2}} \quad\text{...(19).}$$

It is necessary to obtain one of them separately by direct integration.

To obtain the value of J_1 we use confocal coordinates. The level surfaces are

(1) the system of ellipsoids

$$\frac{x_1^2}{p^2+\lambda} + \frac{y_1^2}{\lambda} = 1, \text{ rotated through } \pi \quad\text{.........(20),}$$

(2) the system of hyperboloids

$$\frac{x_1^2}{p^2+\mu} - \frac{y_1^2}{\mu} = 1, \text{ rotated through } \pi \quad \text{.........(21)},$$

and

(3) a series of planes passing through the axis of x which make an angle ϕ with the standard plane. The coordinates are then λ, μ and ϕ. As the problem is one of revolution we only need to consider a section by the plane of xy. Let dS be an element of area in this plane included between the two ellipses characterized by λ and $\lambda + d\lambda$ and the hyperbolas characterized by μ and $\mu + d\mu$. Solving the ellipse and hyperbola equations simultaneously we find that the Cartesian coordinates of the points of intersection are

$$x_1 = \frac{1}{p}\sqrt{(p^2+\lambda)(p^2-\mu)}, \quad y_1 = \frac{1}{p}\sqrt{\lambda\mu}.$$

Let $ABCD$ represent dS magnified.

Let $A = x_1, y_1$, then

$$B = \left(x_1 + \frac{\partial x_1}{\partial \mu}\,d\mu\right), \ \left(y_1 + \frac{\partial y_1}{\partial \mu}\,d\mu\right),$$

$$D = \left(x_1 + \frac{\partial x_1}{\partial \lambda}\,d\lambda\right), \ \left(y_1 + \frac{\partial y_1}{\partial \lambda}\,d\lambda\right),$$

$$AB = \left\{(d\mu)^2\left(\left(\frac{\partial x_1}{\partial \mu}\right)^2 + \left(\frac{\partial y_1}{\partial \mu}\right)^2\right)\right\}^{\frac{1}{2}},$$

$$AD = \left\{(d\lambda)^2\left(\left(\frac{\partial x_1}{\partial \lambda}\right)^2 + \left(\frac{\partial y_1}{\partial \lambda}\right)^2\right)\right\}^{\frac{1}{2}},$$

$$dS = AB\,.\,AD \sin\angle BAD.$$

Let the direction cosines of AD and AB be l, m and l', m' respectively. Then

$$l = \frac{\partial x_1}{\partial \lambda} \div \sqrt{\left(\frac{\partial x_1}{\partial \lambda}\right)^2 + \left(\frac{\partial y_1}{\partial \lambda}\right)^2}, \quad l' = \frac{\partial x_1}{\partial \mu} \div \sqrt{\left(\frac{\partial x_1}{\partial \mu}\right)^2 + \left(\frac{\partial y_1}{\partial \mu}\right)^2}, \text{ etc.,}$$

so that

$$dS = AB\,.\,AD\,(lm' - l'm)$$

$$= d\lambda\,d\mu\left\{\frac{\partial x_1}{\partial \lambda}\frac{\partial y_1}{\partial \mu} - \frac{\partial x_1}{\partial \mu}\frac{\partial y_1}{\partial \lambda}\right\}.$$

From the expressions for x_1, y_1 at intersection

$$\frac{\partial x_1}{\partial \lambda} = \frac{\sqrt{p^2-\mu}}{2p\sqrt{p^2+\lambda}}, \quad \frac{\partial x_1}{\partial \mu} = -\frac{\sqrt{p^2+\lambda}}{2p\sqrt{p^2-\mu}},$$

$$\frac{\partial y_1}{\partial \lambda} = \frac{1}{2p}\sqrt{\frac{\mu}{\lambda}}, \quad \frac{\partial y_1}{\partial \mu} = \frac{1}{2p}\sqrt{\frac{\lambda}{\mu}}.$$

Whence
$$dS = d\lambda\, d\mu\, \frac{\lambda+\mu}{4\sqrt{\mu\lambda(p^2+\lambda)(p^2-\mu)}}.$$

The element of volume $d\tau'$ is obtained by rotating dS about the axis of motion. It is therefore given by

$$\begin{aligned} d\tau' &= 2\pi y' dS \\ &= 2\pi d\lambda\, d\mu\, \frac{\lambda+\mu}{4p\sqrt{(p^2+\lambda)(p^2-\mu)}}. \end{aligned}$$

To calculate the value of $\frac{\partial \phi_1}{\partial x_1}$ we remember that ϕ_1 is constant for any of the confocal ellipsoids and therefore depends only on λ.

Thus
$$\frac{\partial \phi_1}{\partial x_1} = \frac{\partial \phi_1}{\partial \lambda}\frac{\partial \lambda}{\partial x_1}.$$

From the equation $\frac{x_1^2}{p^2+\lambda} + \frac{y_1^2}{\lambda} = 1$,

we find
$$\begin{aligned} \frac{\partial \lambda}{\partial x_1} &= \frac{2x_1}{p^2+\lambda} \div \left(\frac{x_1^2}{(p^2+\lambda)^2} + \frac{y_1^2}{\lambda^2}\right) \\ &= \frac{2\lambda}{p(\lambda+\mu)}\sqrt{(p^2+\lambda)(p^2-\mu)}, \end{aligned}$$

and
$$\begin{aligned} J_1 &= \iiint \left(\frac{\partial \phi_1}{\partial x_1}\right)^2 d\tau_1 \\ &= 2\pi \iint \left(\frac{\partial \phi_1}{\partial \lambda}\right)^2 \frac{\lambda^2\sqrt{(p^2+\lambda)(p^2-\mu)}}{p^3(\lambda+\mu)}\, d\lambda\, d\mu. \end{aligned}$$

The limits of λ are $\lambda = b^2$ to $\lambda = \infty$. With the limits $\mu = 0$ and $\mu = p^2$ we cover half the space, so that

$$J_1 = 4\pi \int_{b^2}^{\infty} d\lambda \left(\frac{\partial \phi_1}{\partial \lambda}\right)^2 \frac{\lambda^2\sqrt{p^2+\lambda}}{p^3} \int_0^{p^2} \frac{\sqrt{p^2-\mu}}{\mu+\lambda}\, d\mu.$$

By changing the variable to x, where $x^2 = p^2 - \mu$, and turning into partial fractions we find

$$\int_0^{p^2} \frac{\sqrt{p^2-\mu}}{\lambda+\mu}\, d\mu = 2p + \sqrt{p^2+\lambda}\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p}.$$

Since
$$\phi_1 = \frac{e}{8\pi p}\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p},$$

$$\frac{\partial\phi_1}{\partial\lambda} = -\frac{e}{8\pi}\div\sqrt{p^2+\lambda},$$

and
$$J_1 = \frac{e^2}{16\pi p^3}\int_{b^2}^{\infty}\left\{\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p} - \frac{2p}{\sqrt{p^2+\lambda}}\right\}d\lambda$$

$$= \frac{e^2}{16\pi p^3}\left|\lambda\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p} - 2p\sqrt{p^2+\lambda}\right|_{b^2}^{\infty}.$$

By expanding

$$\log\frac{\sqrt{p^2+\lambda}+p}{\sqrt{p^2+\lambda}-p} = \log\left(1+\frac{2p}{\sqrt{p^2+\lambda}-p}\right)$$

it can be shown that this vanishes when $\lambda = \infty$. On substituting $p = \sqrt{a^2-b^2}$, $b = R$ and $a = R\div\sqrt{1-\beta^2}$ we find

$$J_1 = \frac{e^2}{16\pi R\beta^3}(1-\beta^2)^{\frac{3}{2}}\left[\frac{2\beta}{1-\beta^2} - \log\frac{1+\beta}{1-\beta}\right] \quad\text{......(22)}.$$

From (19) we have

$$J_1 + J_2 = \frac{e^2}{8\pi R\beta}(1-\beta^2)^{\frac{1}{2}}\log\frac{1+\beta}{1-\beta},$$

whence

$$J_2 = \frac{e^2}{16\pi R\beta^3}(1-\beta^2)^{\frac{1}{2}}\left[-2\beta + (1+\beta^2)\log\frac{1+\beta}{1-\beta}\right] \quad\text{...(23)}.$$

We have seen that the potential (electric) energy is

$$W = \tfrac{1}{2}\left\{(1-\beta^2)^{\frac{1}{2}}J_1 + (1-\beta^2)^{-\frac{1}{2}}J_2\right\},$$

and the kinetic (magnetic) energy is

$$T = \tfrac{1}{2}\beta^2(1-\beta^2)^{-\frac{1}{2}}J_2.$$

The components of the electromagnetic momentum are

$$G_x = -\frac{1}{c}\beta(1-\beta^2)^{-\frac{1}{2}}J_2,$$

$$G_y = -\frac{1}{c}\beta\iiint\frac{\partial\phi_1}{\partial x_1}\frac{\partial\phi_1}{\partial y_1}d\tau_1,$$

and
$$G_z = -\frac{1}{c}\beta\iiint\frac{\partial\phi_1}{\partial x_1}\frac{\partial\phi_1}{\partial z_1}d\tau_1.$$

For a system which is symmetrical about the axis of motion the last two integrals are equal to zero. For, corresponding to any point in the field for which $\frac{\partial \phi_1}{\partial x_1}$ and $\frac{\partial \phi_1}{\partial y_1}$ have certain values, there is another point, which is the reflexion of the first in the plane xOz, for which $\frac{\partial \phi_1}{\partial x_1}$ has the same value but $\frac{\partial \phi_1}{\partial y_1}$ an equal and opposite value. Similar considerations apply to the integral containing $\frac{\partial \phi_1}{\partial z_1}$. If we substitute the values of J_1 and J_2 in the other formulae we get the values of W, T and G_x. For G_x we find

$$G_x = \frac{e^2}{16\pi c R \beta^2} \left[(1+\beta^2) \log \frac{1+\beta}{1-\beta} - 2\beta \right] \text{......(24).}$$

The electromagnetic momentum in the field of the rigid electron was first calculated by Abraham*. The present calculation, which is practically identical with Abraham's, is taken from Lorentz's *Theory of Electrons.*

Electromagnetic Mass.

By expanding the logarithm in the last formula it may be shown that for small values of β the value of G_x reduces to

$$\frac{e^2}{6\pi R c^2} u,$$

where u is the velocity of the moving charge. Again, the corresponding expression for T reduces to $\frac{e^2}{12\pi R c^2} u^2$ when $\beta = \frac{u}{c}$ is small. It will be noticed that these expressions are respectively of the form $m_0 u$ and $\frac{1}{2} m_0 u^2$, where

$$m_0 = \frac{e^2}{6\pi R c^2} \text{ ....(25)}$$

is a constant quantity. The nature of these expressions leads us to a very important result. For they show that the electrified sphere, which we have supposed to be devoid of mass in the ordinary sense, behaves as though it had a mass m_0. Thus *it follows from the principles of electromagnetism alone* that such a charged body, when it moves with the velocity u, carries along with it an amount of electromagnetic momentum $m_0 u$ and of electrokinetic energy $\frac{1}{2}m_0 u^2$. Moreover this momentum and energy remain unaltered so long as the velocity of the body is unaltered. But this behaviour is precisely what characterizes the motion of a so-called material particle whose mass is m_0. It is true that our results would have to be modified for particles moving with velocities comparable with the velocity of light. But they would nevertheless be exact, within the limits of accuracy of measurement, for such relative velocities as have been imparted to any considerable material masses in the universe. On the other hand the differences should be perceptible in the case of the very rapidly moving charged particles emitted by the radioactive substances, and, as we shall see, it is precisely the properties of these particles which have confirmed the results of the electromagnetic theory.

---

* "Prinzipien der Dynamik des Elektrons," *Ann. der Physik*, IV. vol. x. p. 105 (1903).

The idea of electromagnetic inertia, which is due to J. J. Thomson*, is fundamental to the electron theory of matter. For it opens up the possibility that the mass of all matter is nothing else than the electromagnetic mass of the electrons which certainly form part, and perhaps form the whole, of its structure. It obviously opens up the possibility of an electrical foundation for dynamics. This will be considered later.

Our calculations so far have presupposed that the moving charged body possesses and has always possessed a constant velocity in a straight line. A fuller discussion of electromagnetic mass involves the consideration of bodies undergoing acceleration and for such cases the results which we have obtained are not strictly true. This difficulty is one which is peculiar to the electromagnetic theory and arises from the fact that when a charged body is accelerated part of its energy travels off to infinity in the form of electromagnetic radiation.

Longitudinal and Transverse Mass.

We shall, however, see in the next chapter that, provided the acceleration of the body is sufficiently small compared with its velocity, the values for the energy in the field and for the electromagnetic momentum, which we have calculated above, will only differ from the true values by amounts which are exceedingly small. Under these circumstances, which will be stated more precisely later, the formulae which we have obtained for constant velocities may be applied to systems moving with varying velocities. The state of systems which satisfy this condition has been called by Abraham* quasi-stationary.

---

* *Phil. Mag.* vol. XI. p. 229 (1881). The idea that the mass of ordinary matter is of this character on account of the electrons it contains appears to have been first suggested by Larmor (*Phil. Trans.* vol. CLXXXVI. p. 697, 1895).

Assuming that the quasi-stationary condition is satisfied we shall now consider the behaviour of the moving electrical system under the influence of an accelerating force.

There are two cases to consider: (1) when the force is in the direction of motion and (2) when it is perpendicular to it. Any other case may be compounded from these two. We shall suppose the moving charge to be placed in an external field whose action gives rise to the accelerations under consideration. If we consider any infinitely distant surface enclosing the whole electrical system, the Maxwell stresses over it will vanish, so that the force exerted by the external field on the moving charge will be equal to the rate of *diminution* of the electromagnetic momentum of the external field. But the total momentum of the whole system remains constant, so that the external force must be equal to the rate of *increase* of the momentum of the moving charge. Since this is, by hypothesis, massless it follows that the force exerted on the moving charge by the external field is equal to the rate of increase of the electromagnetic momentum of the charge. This equality is clearly a vectorial one and is therefore true for the different components of the force and the momentum independently of one another.

In the case of a force acting in the direction of motion we evidently have

$$F_x = \frac{\partial}{\partial t}(G_x) \doteq \frac{\partial G_x}{\partial \dot{x}} \frac{\partial \dot{x}}{\partial t},$$

since G_x only contains t implicity through $\dot{x}$. Comparing this with the equation $F_x = m\dfrac{\partial \dot{x}}{\partial t}$ we see that the mass for longitudinal accelerations is

$$\frac{\partial G_x}{\partial \dot{x}} = \frac{1}{c}\frac{\partial G}{\partial \beta} \quad \text{........................(26).}$$

---

* *Loc. cit.*

When the force is transverse to the direction of motion the body will begin to describe a circular orbit with constant speed. If r is the instantaneous radius of curvature of the path, the rate of change of the momentum G is always directed along r and, by the principle of the hodograph, is equal per unit time to Gu/r, where u is the instantaneous velocity. Hence

$$F_y = \frac{G}{r} u.$$

But the instantaneous acceleration is also along r and equal to u^2/r, whence it follows that the electromagnetic mass for transverse accelerations is equal to

$$\frac{G}{r} u \div \frac{u^2}{r} = \frac{G}{u} = \frac{1}{c} \frac{G}{\beta} \quad \text{..................(27).}$$

The mass for transverse accelerations is therefore different from the mass for longitudinal accelerations. This difference was first pointed out by Abraham*. It happens, as may easily be verified, that the difference between the longitudinal and the transverse electromagnetic mass becomes vanishingly small for small velocities.

The properties of an electrically charged body which we have been considering have a close analogy in hydrodynamics. Any geometrical figure moving in a fluid sets the surrounding fluid in motion. In the steady state when the figure moves uniformly in a straight line the fluid motion is carried along by the moving figure as though it were rigidly attached to it; when the state of motion changes, waves are set up and part of the energy of the system is radiated away to great distances. We shall see in the next chapter that this also has its counterpart in the electrical case. Confining ourselves to the case of uniform motion, in the steady state it is found† that if the moving figure is intrinsically massless it nevertheless possesses inertia and behaves as though it had a certain mass coefficient which is a function of the mass of the fluid displaced by it. In the case of a massless sphere moving in a perfect fluid this apparent mass is one-half of that of the fluid displaced by the sphere. When a circular cylinder moves at right angles to its length the apparent mass is equal to that of the fluid displaced by the cylinder. In the case of unsymmetrical figures the apparent mass is no longer the same for motions in different directions. The hydrodynamical cases differ from the electrical ones in one important respect; the apparent mass is always independent of the velocity of the moving figure.

---

* *Loc. cit.*

† Lamb's *Hydrodynamics*, pp. 85, 130.

This analogy was very clearly seen by J. J. Thomson in the original memoir* in which he developed the idea of electromagnetic mass. We shall take the liberty of quoting his exact words: "The charged sphere will produce an electric displacement throughout the field; and as the sphere moves, the magnitude of this displacement at any point will vary. Now, according to Maxwell's theory, a variation in the electric displacement produces the same effect as an electric current; and a field in which electric currents exist is a seat of energy; hence the motion of the charged sphere has developed energy, and consequently the charged sphere must experience a resistance as it moves through the dielectric. But as the theory of the variation of the electric displacement does not take into account anything corresponding to resistance in conductors, there can be no dissipation of energy through the medium; hence the resistance cannot be analogous to an ordinary frictional resistance, but must correspond to the resistance theoretically experienced by a solid in moving through a perfect fluid. In other words, it must be equivalent to an increase in the mass of the charged moving sphere."

The Contractile Electron.

We have seen that the determination of the field due to a rigid spherical shell of electrification in motion can be reduced to the determination of the electrostatic potential due to a certain ellipsoid. We have pointed out already that the particular ellipsoid which we have been led to consider as the equivalent fixed system depends upon our supposition that the spherical shell is rigid. Now the negative results which have been obtained in a number of optical experiments on moving systems, instituted largely in order to try to detect relative motion between the system and the luminiferous medium, seem incapable of explanation except on the hypothesis, suggested by FitzGerald†, that, on account of the motion, the matter of the testing system undergoes contraction in the direction of motion. The negative results are immediately accounted for if this contraction is in the ratio of 1 to $\sqrt{1-\beta^2}$, where β is the ratio of the velocity of the system to the velocity of light. The dimensions transverse to the direction of motion are supposed to be unchanged (see Chap. XIII).

---

* *Phil. Mag.* V. vol. II. p. 230 (1881).

† *Nature*, June 16 (1892).

Let us suppose that this contraction affects the electrons as well as the material as a whole. The resolution of the problem is due to Lorentz*. It is in reality much simpler than that of the rigid spherical shell. For the shell which was spherical when at rest becomes an oblate spheroid when in motion, the polar axis coinciding with the direction of motion and being equal to $R\sqrt{1-\beta^2}$, where R is the original radius of the shell. The equatorial radius is unchanged and equal to R. The question now arises as to what is the corresponding fixed system S_1. This will be obtained if we multiply all lengths parallel to the direction of motion by $1/\sqrt{1-\beta^2}$, leaving the perpendicular directions unchanged. Thus the corresponding fixed system is simply a sphere of radius R. The potential ϕ_1 is symmetrical in the distorted space of the fixed system and equal to $e/4\pi r_1$, where r_1 is the distance from the centre of the sphere in this system. The electrostatic energy is thus

$$= \tfrac{1}{2}e\phi_0 = \frac{e^2}{8\pi R}$$

$$= \tfrac{1}{2}\iiint\left\{\left(\frac{\partial\phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\} d\tau_1.$$

But everything is symmetrical about the centre of the sphere in the fixed system, so that

$$\iiint\left(\frac{\partial\phi_1}{\partial x_1}\right)^2 d\tau_1 = \iiint\left(\frac{\partial\phi_1}{\partial y_1}\right)^2 d\tau_1 = \iiint\left(\frac{\partial\phi_1}{\partial z_1}\right)^2 d\tau_1,$$

and $G_x = \frac{1}{c}\beta(1-\beta^2)^{-\frac{1}{2}}\iiint\left\{\left(\frac{\partial\phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\} d\tau_1$

$$= \tfrac{2}{3}\frac{1}{c}\beta(1-\beta^2)^{-\frac{1}{2}}\iiint\left\{\left(\frac{\partial\phi_1}{\partial x_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial y_1}\right)^2 + \left(\frac{\partial\phi_1}{\partial z_1}\right)^2\right\} d\tau_1$$

$$= \frac{e^2\beta}{6\pi cR\sqrt{1-\beta^2}},$$

also $G_y = G_z = 0$, as before.

---

* *Theory of Electrons*, p. 210.

For longitudinal accelerations we have

$$\begin{aligned} F_x = m_1 \frac{\partial^2 x_1}{\partial t^2} &= \frac{\partial G_x}{\partial t} \\ &= \frac{\partial G_x}{\partial \beta}\frac{\partial \beta}{\partial t} = \frac{e^2}{6\pi Rc}\left\{(1-\beta^2)^{-\frac{1}{2}} + \beta^2 (1-\beta^2)^{-\frac{3}{2}}\right\} \times \frac{\partial \beta}{\partial t} \\ &= \frac{e^2}{6\pi Rc^2 (1-\beta^2)} \frac{\partial^2 x}{\partial t^2}, \end{aligned}$$

so that the longitudinal mass

$$m_1 = \frac{e^2}{6\pi Rc^2 (1-\beta^2)^{\frac{3}{2}}} = \frac{m_0}{(1-\beta^2)} \quad \text{............(28),}$$

where m_0 is the value of the mass for small velocities.

For transverse accelerations

$$\left.\begin{aligned} m_2 = \frac{G_x}{u} = \frac{G_u}{c\beta} &= \frac{e^2}{6\pi Rc^2 (1-\beta^2)^{\frac{1}{2}}} \\ &= \frac{m_0}{(1-\beta^2)^{\frac{1}{2}}} \end{aligned}\right\} \text{.........(29).}$$

Thus the ratio of the longitudinal to the transverse mass is

$$\frac{m_1}{m_2} = (1-\beta^2)^{-1} \quad \text{.....................(30).}$$

These formulae are all simpler than the corresponding ones for the rigid electron.

The Experimental Evidence.

We saw in Chapter I that the experiments of Thomson, Wiechert and others had established the existence of particles whose charge per unit mass e/m had a value some 1800 times as great as that for an atom of hydrogen in electrolysis. The smallness of the mass of these particles together with the relative largeness of their charge suggested them as a likely field in which to look for experimental evidence of the existence of electromagnetic mass. A means for the detection of the latter is furnished by the fact that electromagnetic mass is a function of the velocity of the moving charge, whereas the ordinary mass of the Newtonian scheme is assumed to be independent of the velocity of the particle.

It is true that the alteration in the mass is small except when the velocity is comparable with the velocity of light; but it happens, fortunately, that in the β-rays from radioactive substances we have moving electrons whose velocities vary widely but extend almost up to the velocity of light.

The first successful experiments in this direction were made by Kaufmann*. Working with the β-rays from radium bromide he made use of a device similar to that of crossed prisms in experiments on dispersion. His apparatus is shown in Fig. 30 *a*. A speck of radium bromide was placed at the point O immediately below the parallel plate condenser of which P_1P_2 represents a section by the plane of the paper. Immediately above P_1P_2 was a minute hole D in a thick metal plate. The line OD was vertical, and in the plane midway between the condenser plates. At some distance above D was a horizontal photographic plate. A suitable difference of electrostatic potential was established between the plates of the condenser, and the whole system was placed between coils designed to produce a uniform magnetic field and lying in planes at right angles to that of the paper. Thus the particles were acted upon, during their passage between the plates, by a horizontal electric force lying in the plane of the paper. If this field alone were operative the particles passing through the points O and D would pursue parabolic paths when between the condenser plates P_1 P_2, the parabolas lying in the plane of the paper. After escaping from the plate condenser the subsequent path would be rectilinear and along the tangent to the parabola at the point of escape. Thus under

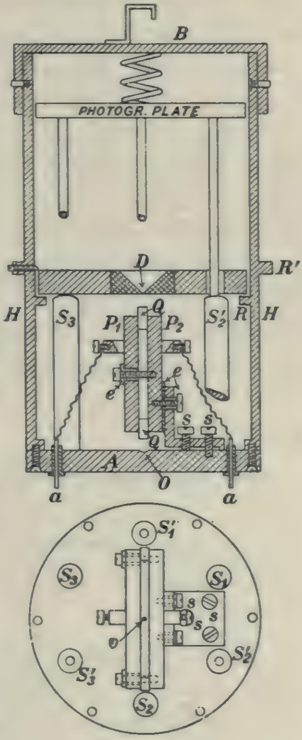


Fig. 30 *a*.

---

* *Gött. Nachr.* 1901, Heft 1; 1902, Heft 5; 1903, Heft 3; *Phys. Zeits.* p. 55, 1902; *Ann. der Phys.* IV. vol. XIX. p. 487, 1906 (complete account).

the influence of the electrostatic field alone, a group of particles all having the same velocity would be deflected to the right or to the left according to the direction of the field. The spot would be shifted from the centre of the photographic plate. Since the deflection depends upon the velocity, if the group consisted of particles having different velocities the spot would be spread out into a line lying in the plane of the paper. Each point on this line would correspond to a particular velocity.

The lines of force of the magnetic field run from left to right in the plane of the paper, so that if it alone were operative it would cause the particles which pass through O and D to describe circles in a plane perpendicular to the plane of the paper. Owing to the magnetic field, then, the spot would be displaced in the plane perpendicular to the plane of the paper, if all the particles had the same velocity, and would be drawn out into a line in this plane, if the particles had different velocities. When both fields are operative at once we should expect to get a curved line on the photographic plate, each point of which represents the point of impact of particles having a certain velocity. The position of each point gives us, of course, the magnetic and electrostatic deflection of a particle with a definite but unknown velocity. It remains to be seen how we may deduce from the measured displacements the value of the mass as a function of the velocity of the moving particle.

For the sake of simplicity we shall suppose a uniform electric field to extend, from left to right, all the way from O to D and then to cease absolutely. The magnetic field H is uniform and parallel to this electric field all the way from O to the photographic plate P.

Taking O as the origin, let the coordinates of D be $0, 0, z$, and those of points on the plate P be x, y, z'. Let the axis of y lie in the plane of the paper, Y denoting the electric intensity. Assuming that the deflections may be treated as small, we have, so far as the motion in the plane of the paper is concerned,

$$m\frac{\partial^2 z}{\partial t^2} = 0 \text{ and } m\frac{\partial^2 y}{\partial t^2} = Ye, \text{ from } O \text{ to } D,$$

and

$$m\frac{\partial^2 y}{\partial t^2} = 0, \text{ from } D \text{ to } P.$$

Thus $$\frac{\partial z}{\partial t} = \text{const.} = w_0,$$

where w_0 is the component along OD of the velocity of projection,

and $$z = w_0 t + \text{const.} = w_0 t$$

if $t=0$ when the particle leaves O.

Hence $t = z/w_0$ and

$$Ye = m\frac{\partial^2 y}{\partial t^2} = m w_0^2 \frac{\partial^2 y}{\partial z^2}.$$

Thus $$m w_0^2 y = \tfrac{1}{2} Yez^2 + Az + B.$$

Since $y=0$ when $z=0$ and also when $z=z_1$,

$$B = 0 \text{ and } A = -\tfrac{1}{2} Yez_1.$$

Thus $$m w_0^2 y = \tfrac{1}{2} Yez(z-z_1),$$

$$\frac{\partial y}{\partial z} = \frac{1}{m w_0^2}(Yez - \tfrac{1}{2} Yez_1),$$

and $$\left(\frac{\partial y}{\partial z}\right)_{z=z_1} = \frac{1}{m w_0^2}(\tfrac{1}{2} Yez_1) = \tan\theta,$$

where θ is the angle which the tangent to the parabolic path at D makes with OD. Since the subsequent path is rectilinear the displacement y' at the photographic plate is

$$y' = (z'-z_1)\tan\theta = \frac{1}{2 m w_0^2} Yez_1(z'-z_1).$$

The projection of the path on the plane of xOz is controlled by the magnetic field and is a circle passing through O and D to the degree of approximation of this calculation. Let r be the radius of this circle and $S_1D_1P_1$ (Fig. 30 *b*) the tangent to it at the point S_1 which is symmetrical with respect to O and D. D_1 and P_1 are the intersections with horizontal planes through D and P respectively. $P'(x', 0, z')$ is the intersection of the projected trajectory on the photographic plate. S' is the intersection of a vertical line $P'S'$ and a horizontal line S_1SS'. S is the mid-point of OD. Then

$$P_1P = S_1S = \frac{OS^2}{2r - SS_1} = \frac{z_1^2}{8r},$$

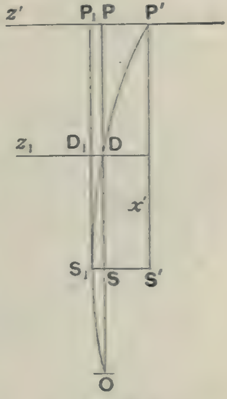


Fig. 30 *b*.

neglecting SS_1 compared with r, and

$$P_1P' = S_1S' = \frac{S_1P_1^2}{2r - P_1P'} = \frac{(z' - \frac{1}{2}z_1)^2}{2r},$$

neglecting P_1P' compared with r. Thus

$$x' = PP' = P_1P' - P_1P = \frac{z'(z' - z_1)}{2r}.$$

The radius of curvature of the circular path is determined by the balancing of the centrifugal force by the force exerted by the magnetic field on the moving particle. Thus

$$\frac{mw_0^2}{r} = Hew_0,$$

when r is sufficiently large. Hence

$$x' = \frac{Hez'}{2mw_0}(z' - z_1).$$

Comparing this with the electrostatic deflection y', we have

$$w_0 = \frac{x'}{y'}\frac{Y}{H}\frac{z_1}{z'} \quad \text{........................(31)},$$

and eliminating w_0

$$\frac{m}{e} = \frac{y'}{x'^2}\frac{H^2}{Y}\frac{z'^2(z' - z_1)}{2z_1} \quad \text{..................(32)}.$$

These equations are only to be taken as illustrating the principle of the method. Owing to the fact that the electric field does not extend all the way from the point O to the point D and that the electric and magnetic fields are not quite uniform, the actual treatment of the experimental data is somewhat different. For these refinements the reader must consult the original paper.

The experiments are carried out with certain values of H and Y. We see from equation (31) that if we take the undeflected spot as origin and the axes of magnetic and electric displacement respectively as the x and y axes, then the ratio of the coordinates of any point on the curve on the plates determines a certain value of the velocity w_0 of the particles. Substituting the same values of x' and y' in equation (32), we obtain the value of m/e possessed by the particle whose initial velocity had the particular value previously found. We can thus find the value of m/e for all the velocities present among the β-ray particles. Since it is highly probable from other considerations that e is constant we are in this way able to find m as a function of w_0.

From the results of Kaufmann's experiments it was immediately clear that the mass of these particles increased with increasing velocity and that the variation was in good agreement with the formulae which had been developed on the supposition that the whole of the mass was of electromagnetic origin. A careful re-examination of his plates led Kaufmann to the conclusion that his results agreed better with the formula of Abraham for the rigid electron than with that of Lorentz for the contractile one. The graphs of the two formulae are however not very different for the range of velocities embraced by the β-rays used, and it seemed at the time very questionable whether the experiments did not agree with the results predicted by both formulae within the limits of experimental error. It is to be borne in mind that this experiment tests only the transverse mass and tells nothing about the mass for longitudinal accelerations.

Bucherer's Experiment.

A very ingenious experiment to test the different theories of the constitution of the electron has been carried out by Bucherer*.

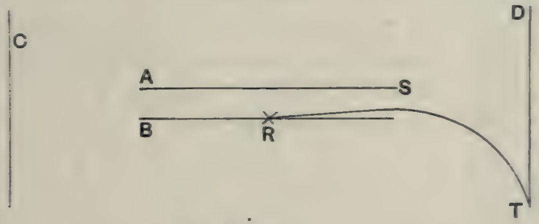


Fig. 31.

A speck of radium fluoride, which contains more radium per gramme than any other available compound of radium, is placed at the point R (Fig. 31), which lies at the centre of the lower plate of a circular parallel plate condenser AB. The distance between A and B is very small compared with the dimensions of the plates. The plates A and B are maintained at a suitable difference of potential and are placed, so that their planes are horizontal, in the centre of a vertical cylinder. CD is a section of the walls of the cylinder by

---

* *Ann. der Phys.* IV. vol. xxviii. p. 513 (1909).

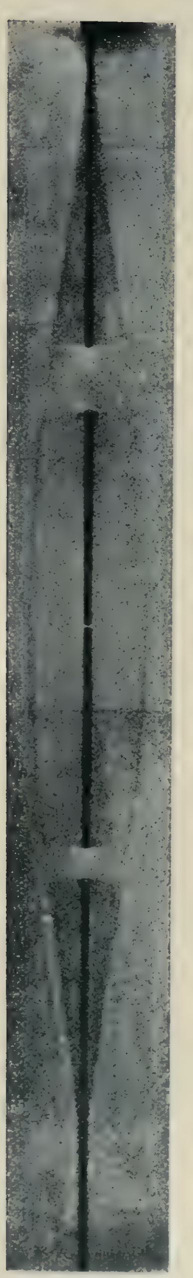


Fig. 32.

the plane of the paper. A photographic film extends all the way round the inside of the cylinder. The whole apparatus is exhausted and is placed in a uniform magnetic field H perpendicular to the plane of the paper.

To simplify the discussion of the experiment we shall suppose the distance between A and B to be so small that it is quite negligible. In the absence of the electric and magnetic fields the β-rays travel in straight lines all the way from R and their trace on the photographic film is a circle in the plane of the condenser. This circle becomes a straight line, of course, when the film is unrolled. When the fields are applied the paths become more complex. Between the plates they are still horizontal straight lines, but after escaping they describe circular spirals with their axes along the direction of the magnetic force. The rays will only be able to escape from the plates provided the downward pull of the magnetic force is equal to the upward pull of the electric field. If we consider a particle starting out with velocity w in a direction making an angle α with that of the magnetic field, the condition for compensation is

$$Xe = Hew \sin \alpha,$$

where X is the electric intensity between the plates. Thus for the particles to escape from the condenser

$$\beta = \frac{w}{c} = \frac{X}{Hc \sin \alpha}.$$

In Bucherer's experiments this formula was tested by taking X and H so that $X/Hc = \frac{1}{2}$. Under these circumstances

$$\sin \alpha = 1/2\beta.$$

After leaving the plates the particles follow a spiral path of which ST may be regarded as the projection on the plane of the figure. Consider the position of the point T as a function of the angle α. When w is equal to the velocity of light, β has the maximum value unity. For particles moving with this speed the theoretical magnetic deflection is zero on account of the electromagnetic mass being infinite. This follows from the formula for the radius of curvature $r = mw/He$. We should therefore expect the trace on the film to coincide with the undeflected trace (i.e. the rectilinear trace obtained in the absence of the fields) when

$$\sin\alpha = \tfrac{1}{2} \text{ or } \alpha = 30^\circ \text{ or } 150^\circ.$$

This was immediately verified. The maximum deflection occurs when $\alpha = \pi/2$ or $\beta = \frac{1}{2}$ with the electric and magnetic fields used. The form of the experimental trace is shown in Fig. 32, the very dark line being the undeflected position. None of the rays escape from the plate except when the value of α lies between 30° and 150°.

The values of the maximum deflection when $\alpha = \pi/2$ enable us to distinguish between the different formulae for the mass of an electron. Since these particular particles always move at right angles to the magnetic field, their paths are circles. Let r be the radius, let a be the difference between the radii of the cylinder CD and the plates of the condenser AB and let z be the perpendicular deflection from the undeviated trace. Then

$$z = a^2/(2r - z),$$

and therefore

$$r = \tfrac{1}{2}\left(z + \frac{a^2}{z}\right).$$

Thus

$$m = \frac{Her}{w} = \frac{He}{2w}\left(z + \frac{a^2}{z}\right).$$

If m_0 is the mass of the moving electric charge for zero velocity ($w = 0$), then for the contractile electron suggested by Lorentz, we have

$$e/m_0 = \frac{e}{m}(1 - \beta^2)^{-\frac{1}{2}},$$

and substituting the value of m given above we find

$$\frac{e}{m_0} = \frac{2cz}{H(a^2 + z^2)} \frac{\beta}{\sqrt{1 - \beta^2}} = \frac{2cz}{H(a^2 + z^2)} \tan(\sin^{-1}\beta) \ldots (33).$$

With Abraham's rigid electron, on the other hand,

$$\frac{e}{m_0} = \frac{2cz}{H(a^2+z^2)} \left\{ \frac{3}{4\beta} \frac{2\delta - \tanh 2\delta}{\tanh 2\delta} \right\} \quad\text{............(34)},$$

where $\beta = \tanh \delta$.

The relative constancy of the values of e/m_0 calculated from the experimental deflections by means of equations (33) and (34) respectively should enable us to distinguish between the two formulae. The data yielded by the experiments are given in the accompanying table:

$\beta = \frac{w}{c}$	H	z (mms.)	$\frac{e}{m_0}$ (10^{-7}) (Lorentz)	$\frac{e}{m_0}$ (10^{-7}) (Abraham)
·3173	104·55	16·37	1·752	1·726
·3787	115·78	14·45	1·761	1·733
·4281	127·37	13·50	1·760	1·723
·5154	127·55	10·18	1·763	1·706
·5154	127·55	10·35	—	—
·6870	127·55	6·23	1·767	1·642

In the case of the Lorentz electron the value of e/m_0 is constant within the limits of experimental error, whereas in the case of the rigid electron the deviation is much greater than can be accounted for in this way.

These experiments appear to dispose effectually of the rigid electron and they may be regarded as making it reasonably certain that Thomson's corpuscles are devoid of mass except such as is due to the charge that they carry. For this reason we shall always refer to them in the sequel as negative electrons.

We shall find later on that the relation between m and m_0 characteristic of the Lorentz contractile electron is true of all electrical systems according to the principle of relativity. Bucherer's experiment may therefore be regarded as evidence in favour of that principle.

CHAPTER XII

CHARGE MOVING WITH VARIABLE VELOCITY

IN the case of a charged body moving with variable velocity the field is no longer carried along as though it were rigidly attached to the moving system. That is a state of things which is characteristic of uniform rectilinear motion. In order to determine the state of the field in other cases our only recourse is to evaluate the potentials

$$\phi = \frac{1}{4\pi}\iiint \frac{\rho'}{r}\,d\tau \quad \text{and} \quad U = \frac{1}{4\pi}\iiint \frac{(V\rho)'}{r}\,d\tau$$

by direct integration. We shall see how this may be done in the case of a point charge, that is to say of a charged body all of whose dimensions are negligible compared with the other dimensions entering into the problem.

Let us seek the values of the potentials at the point P (Fig. 33) at a certain instant, t. It will be remembered that the values ρ' and $(V\rho)'$ which enter into the integration are not the values at the element $d\tau$ at the instant t for which we are seeking the potentials at P but at a certain instant, say $t - \theta$, which differs from t by the time which is required for a disturbance travelling with the velocity of radiation to pass from the element of volume $d\tau$ to the point P. Clearly $r = c\theta$, if r is the distance from $d\tau$ to P and c is the velocity of light. We shall suppose the coordinates ξ, η, ζ of the moving charged body to be given explicitly as a function of the time $t - \theta$ at which the disturbance leaves it. If δ denotes partial differentiation when t and θ are considered as independent variables, we have

$$\frac{\delta\xi}{\delta\theta} = \frac{d\xi}{d(t-\theta)} \cdot \frac{\partial(t-\theta)}{\partial\theta} = -\frac{d\xi}{d(t-\theta)} = -\frac{\delta\xi}{\delta t} = -\dot{\xi},$$

$$\frac{\delta\dot{\xi}}{\delta\theta} = \frac{d\dot{\xi}}{d(t-\theta)} \cdot \frac{\partial(t-\theta)}{\partial\theta} = -\frac{d\dot{\xi}}{d(t-\theta)} = -\frac{\delta\dot{\xi}}{\delta t} = -\ddot{\xi},$$

with similar equations involving η, $\dot{\eta}$, $\ddot{\eta}$ and ζ, $\dot{\zeta}$, $\ddot{\zeta}$.

At the outset we are face to face with a new difficulty which arises from the fact that the charge $\rho d\tau$ which enters into the integrals expressing the potentials is not all present in the element of volume at the same instant. Thus the charge $\rho d\tau$ which occurs in this element of the integral is not equal to the true charge which would occur in this element if it were at rest. Consider the truncated cone $TOSR$ whose apex is at P and which is terminated by the spherical surfaces $AOSC$ and $BTRD$ whose radii are $c\theta$ and $c(\theta + d\theta)$ respectively. The sphere AC represents the instantaneous position of contributing charges at the instant $t - \theta$ and the sphere BD the instantaneous position at $t - \theta - d\theta$.

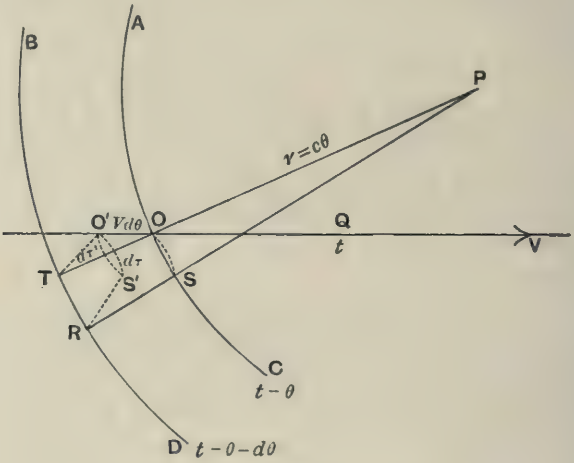


Fig. 33.

Thus the part of the charge which contributes to the potential at P from the front-end OS of the element of volume $d\tau$ is present there at a later time than that which contributes from the back-end TR. All the charges are, however, present in the displaced element $d\tau' = O'S'RT$ at the one instant $t - \theta - d\theta$ provided $OO' = Vd\theta$. Hence the true charge de which is effective in the element $d\tau$ is, if λ is the angle between r and V,

$$de = \rho d\tau' = \rho d\tau\,(1 - \beta \cos \lambda)^*.$$

---

* This result is due to Wiechert, *Arch. Néerl.*, (2) vol. v. p. 549 (1900). Cf. also A. Liénard, *L'Éclairage électrique*, vol. xvi. pp. 5, 53, 106 (1898).

For, since $d\tau$ and $d\tau'$ have the same base,

$$\frac{d\tau'}{d\tau} = \frac{OT - Vd\theta\cos\lambda}{OT} = 1 - \frac{Vd\theta\cos\lambda}{c\,d\theta}.$$

We therefore obtain the following expressions for the potentials:

$$\phi = \frac{1}{4\pi}\iiint\left(\frac{de}{r(1-\beta\cos\lambda)}\right)', \quad U = \frac{1}{4\pi c}\iiint\left(\frac{Vde}{r(1-\beta\cos\lambda)}\right)' \ldots(1),$$

and for a point charge e

$$\phi = \frac{e'}{4\pi r(1-\beta\cos\lambda)}, \quad U = \frac{(Ve)'}{4\pi rc(1-\beta\cos\lambda)} \ldots\ldots(2).$$

The dashes denote that the values are not the instantaneous ones but those at a time $\theta = r/c$ previous. We have assumed in this investigation that the effect of any real change in ρ may be regarded as negligibly small during the time $d\theta$.

Having obtained ϕ and U, the electric and magnetic vectors are now to be obtained from the equations

$$E = -\operatorname{grad}\phi - \frac{1}{c}\frac{\partial U}{\partial t},$$

and $$H = \operatorname{rot} U.$$

The differentiations which enter into these equations are at the point P fixed with respect to the axes of reference. The independent variables are now x, y, z and t. The potentials involve x, y, z partly through r and partly through θ, which is a function of x, y, z when t is fixed and involves ξ, η, ζ which are functions of $(t-\theta)$. They also depend on t both directly through ξ, η, ζ and also indirectly through θ. We have

$$r = c\theta = \sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2},$$

$$\cos\lambda = \frac{1}{Vr}\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\},$$

$$r(1-\beta\cos\lambda) = r - \frac{1}{c}\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}.$$

Let us seek the value of

$$E_x = -\frac{\partial\phi}{\partial x} - \frac{1}{c}\frac{\partial U_x}{\partial t}.$$

We have

$$\phi = \frac{e}{4\pi} \frac{1}{c\theta - \frac{1}{c}\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}},$$

$$-\frac{\partial\phi}{\partial x}$$

$$= \frac{ec}{4\pi} \frac{c^2 \frac{\partial\theta}{\partial x} - \dot{\xi} + \dot{\xi}\frac{\partial\xi}{\partial x} + \dot{\eta}\frac{\partial\eta}{\partial x} + \dot{\zeta}\frac{\partial\zeta}{\partial x} - (x-\xi)\frac{\partial\dot{\xi}}{\partial x} - (y-\eta)\frac{\partial\dot{\eta}}{\partial x} - (z-\zeta)\frac{\partial\dot{\zeta}}{\partial x}}{[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}]^2}.$$

But $$\frac{\partial\xi}{\partial x} = \frac{d\xi}{d(t-\theta)} \frac{\partial(t-\theta)}{\partial\theta} \frac{\partial\theta}{\partial x} = -\dot{\xi}\frac{\partial\theta}{\partial x}.$$

Similarly

$$\frac{\partial\eta}{\partial x} = -\dot{\eta}\frac{\partial\theta}{\partial x}, \quad \frac{\partial\zeta}{\partial x} = -\dot{\zeta}\frac{\partial\theta}{\partial x}, \quad \frac{\partial\dot{\xi}}{\partial x} = -\ddot{\xi}\frac{\partial\theta}{\partial x},$$

$$\frac{\partial\dot{\eta}}{\partial x} = -\ddot{\eta}\frac{\partial\theta}{\partial x} \quad \text{and} \quad \frac{\partial\dot{\zeta}}{\partial x} = -\ddot{\zeta}\frac{\partial\theta}{\partial x}.$$

Hence

$$-\frac{\partial\phi}{\partial x}$$

$$= \frac{ec}{4\pi} \frac{-\dot{\xi} + \{c^2 - (\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2) + (x-\xi)\ddot{\xi} + (y-\eta)\ddot{\eta} + (z-\zeta)\ddot{\zeta}\}\frac{\partial\theta}{\partial x}}{[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}]^2},$$

and

$$c\frac{\partial\theta}{\partial x} = \frac{1}{r}\left\{x - \xi - (x-\xi)\frac{\partial\xi}{\partial x} - (y-\eta)\frac{\partial\eta}{\partial x} - (z-\zeta)\frac{\partial\zeta}{\partial x}\right\},$$

or $$\frac{\partial\theta}{\partial x} = \frac{x-\xi}{c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}}.$$

Thus

$$-\frac{\partial\phi}{\partial x} = \frac{ec}{4\pi} \frac{\begin{aligned}[c^2\{(x-\xi) - \theta\dot{\xi}\} + \dot{\xi}\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\} \\ - (x-\xi)\{\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2\} \\ + (x-\xi)\{(x-\xi)\ddot{\xi} + (y-\eta)\ddot{\eta} + (z-\zeta)\ddot{\zeta}\}]\end{aligned}}{(c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\})^3}.$$

The expressions for $-\frac{\partial\phi}{\partial y}$ and $-\frac{\partial\phi}{\partial z}$ may be written down from inspection, on interchanging the axes.

$$U_x = \frac{e\dot{\xi}}{4\pi\left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right]},$$

$$\begin{aligned}-\frac{1}{c}\frac{\partial U_x}{\partial t} = \frac{e}{4\pi c}\Bigg\{ & -\frac{\partial\dot{\xi}}{\partial t}\left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right] + \dot{\xi}c^2\frac{\partial\theta}{\partial t} \\ & + \dot{\xi}\left[\dot{\xi}\frac{\partial\xi}{\partial t} + \dot{\eta}\frac{\partial\eta}{\partial t} + \dot{\zeta}\frac{\partial\zeta}{\partial t} - \left\{(x-\xi)\frac{\partial\dot{\xi}}{\partial t} + (y-\eta)\frac{\partial\dot{\eta}}{\partial t}\right.\right. \\ & \left.\left. + (z-\zeta)\frac{\partial\dot{\zeta}}{\partial t}\right\}\right]\Bigg\} \\ & \div \left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right]^2,\end{aligned}$$

$$\frac{\partial\xi}{\partial t} = \frac{d\xi}{d(t-\theta)}\frac{\partial(t-\theta)}{\partial t} + \frac{d\xi}{d(t-\theta)}\frac{\partial(t-\theta)}{\partial\theta}\frac{\partial\theta}{\partial t} = \dot{\xi}\left(1 - \frac{\partial\theta}{\partial t}\right).$$

Similarly

$$\frac{\partial}{\partial t}(\eta, \zeta, \dot{\xi}, \dot{\eta}, \dot{\zeta}) = \left(1 - \frac{\partial\theta}{\partial t}\right)(\dot{\eta}, \dot{\zeta}, \ddot{\xi}, \ddot{\eta}, \ddot{\zeta}),$$

$$\begin{aligned}c\frac{\partial\theta}{\partial t} &= -\frac{1}{r}\left\{(x-\xi)\frac{\partial\xi}{\partial t} + (y-\eta)\frac{\partial\eta}{\partial t} + (z-\zeta)\frac{\partial\zeta}{\partial t}\right\} \\ &= -\frac{1}{r}\left(1 - \frac{\partial\theta}{\partial t}\right)\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\},\end{aligned}$$

$$\frac{\partial\theta}{\partial t} = -\frac{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}}{c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}},$$

$$1 - \frac{\partial\theta}{\partial t} = \frac{c^2\theta}{c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}}.$$

Hence

$$\begin{aligned}-\frac{1}{c}\frac{\partial U_x}{\partial t} = \frac{ec}{4\pi}\Big\{ & -\ddot{\xi}\theta\left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right] \\ & - \dot{\xi}\{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\} \\ & + \dot{\xi}\theta\left[\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2 - \{(x-\xi)\ddot{\xi} + (y-\eta)\ddot{\eta} + (z-\zeta)\ddot{\zeta}\}\right]\Big\} \\ & \div \left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right]^3,\end{aligned}$$

and

$$\begin{aligned}E_x = {} & \frac{ec}{4\pi\left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right]^3} \\ & \times \Big[(x - \xi - \theta\dot{\xi})\{c^2 - (\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2)\} \\ & + (x - \xi - \theta\dot{\xi})\{(x-\xi)\ddot{\xi} + (y-\eta)\ddot{\eta} + (z-\zeta)\ddot{\zeta}\} \\ & - \ddot{\xi}\theta\left[c^2\theta - \{(x-\xi)\dot{\xi} + (y-\eta)\dot{\eta} + (z-\zeta)\dot{\zeta}\}\right]\Big].\end{aligned}$$

The expressions for E_y and E_z may be written down by interchanging the axes.

It will be observed that of the three terms in the expression for E_x the first does not involve the acceleration $\ddot{\xi}, \ddot{\eta}, \ddot{\zeta}$. Let us consider this part of the electric intensity separately. Its components are proportional to

$$(x-\xi-\theta\dot{\xi}),\quad (y-\eta-\theta\dot{\eta}) \text{ and } (z-\zeta-\theta\dot{\zeta})$$

respectively. The resultant of this part of the electric intensity is therefore proportional to and directed along O_1P, where O_1 is the position which the moving charge would occupy if it continued from the instant $t-\theta$ to move uniformly during the following interval θ with the velocity $\dot{\xi}, \dot{\eta}, \dot{\zeta}$ which it had at that instant. We may therefore write the resultant E_1 of the part of the electric intensity which does not depend upon the acceleration in the form

$$E_1 = \frac{e(1-\beta^2)}{4\pi r^3(1-\beta\cos\lambda)^3} O_1P \text{ ....(3).}$$

If we work out the value of the magnetic intensity from the expression $H = \text{rot } U$ we find that there is a part of that also which is independent of the acceleration. Denoting this by H_1 we find

$$H_1 = \beta E_1 \sin\lambda_1 \text{ ....(4),}$$

where λ_1 is the angle between O_1P and the direction of V. H_1 is tangential to the circle passing through P in the plane perpendicular to V and whose centre is on the direction of V.

In the case of a particle which moves with a uniform velocity, the expressions just given will represent the whole of the electric and magnetic intensities. In that case it is convenient to express $E_1 = E$ in terms of the instantaneous radius r_1 and the angle it makes with the direction of motion, rather than in terms of r and λ. The change is easily made. We have

$$r_1/\sin\lambda = r/\sin\lambda_1$$

and since

$$OO_1 = V\theta = \beta r,$$

$$\begin{aligned} r_1^2 &= r^2 + \beta^2 r^2 - 2\beta r^2\cos\lambda \\ &= r^2(1-\beta\cos\lambda)^2 + \beta^2 r^2\sin^2\lambda \\ &= r^2(1-\beta\cos\lambda)^2 + \beta^2 r_1^2\sin^2\lambda_1; \end{aligned}$$

hence $$r(1-\beta\cos\lambda)=r_1(1-\beta^2\sin^2\lambda_1)^{\frac{1}{2}},$$

and $$E=E_1=\frac{e(1-\beta^2)}{4\pi r_1^2(1-\beta^2\sin^2\lambda_1)^{\frac{3}{2}}}\quad\ldots\ldots\ldots\ldots\ldots(5),$$

$$H=H_1=\beta E_1\sin\lambda_1.$$

These formulae were first given by Heaviside*.

In the case of uniform rectilinear motion the resultant electric intensity at any point P is directed along the radius from P to the instantaneous position O_1 of the moving charge. It is a maximum where $\sin\lambda_1$ is a maximum, that is, in the equatorial plane, and is a minimum where $\sin\lambda_1=0$, that is, along the polar axis. The same thing is true of the magnetic force, which however vanishes in the neighbourhood of the polar axis. This state of affairs has been described by J. J. Thomson in the statement that the lines of force due to an electric charge in motion tend to concentrate in the equatorial plane. In the case when the velocity becomes equal to the velocity of light the concentration is complete. Since β is then equal to unity the force vanishes at every point outside of the equatorial plane. It will be observed that this part of the force varies inversely as the square of the distance from the moving charge and is therefore inappreciable at big distances.

Acceleration.

Turning to E_2, the part of E which involves the acceleration, we see that there are two terms in this, and their components are proportional respectively to

$$(x-\xi-\theta\dot{\xi}),\quad(y-\eta-\theta\dot{\eta}),\quad(z-\zeta-\theta\dot{\zeta})$$

and $\ddot{\xi}$, $\ddot{\eta}$, $\ddot{\zeta}$. Thus E_2 may be regarded as being made up of a part which is directed along O_1P together with a component parallel to the acceleration. If we call the acceleration $(\ddot{\xi},\ddot{\eta},\ddot{\zeta})=\overline{\Gamma}$ this result may be written

$$E_2=\frac{e\Gamma\cos\mu}{4\pi c^2r^2(1-\beta\cos\lambda)^3}\,\overline{O_1P}-\frac{e}{4\pi c^2r(1-\beta\cos\lambda)^2}\,\overline{\Gamma}\ldots(6),$$

where μ is the angle between r and $\overline{\Gamma}$.

E_2 obviously lies in the plane of the radius O_1P and the acceleration. It is also at right angles to the radius $r = OP$. For if we multiply each component of E_2 in turn by the corresponding direction cosine

---

* *Electrician*, Dec. 7, 1888, p. 148; *Phil. Mag.* V. vol. XXVII. p. 332 (1889).

$$\frac{x-\xi}{r},\ \frac{y-\eta}{r}\ \text{or}\ \frac{z-\zeta}{r}$$

of OP, we find, for the numerator of the sum,

$$\begin{aligned}&\{(x-\xi)\ddot{\xi}+(y-\eta)\ddot{\eta}+(z-\zeta)\ddot{\zeta}\}\{(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2\\&\qquad\qquad-\theta[(x-\xi)\dot{\xi}+(y-\eta)\dot{\eta}+(z-\zeta)\dot{\zeta}]\}\\&\quad-[c^2\theta-\{(x-\xi)\dot{\xi}+(y-\eta)\dot{\eta}+(z-\zeta)\dot{\zeta}\}]\\&\qquad\qquad\times\theta\{(x-\xi)\ddot{\xi}+(y-\eta)\ddot{\eta}+(z-\zeta)\ddot{\zeta}\}\\&\quad=[(x-\xi)\ddot{\xi}+(y-\eta)\ddot{\eta}+(z-\zeta)\ddot{\zeta}]\\&\qquad\qquad\times[(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2-c^2\theta^2]\\&\quad=0.\end{aligned}$$

Thus E_2 is at right angles to OP and in the plane containing O_1P and the acceleration.

If we work out the value H_2 of the part of the magnetic intensity H which involves the acceleration we find that this is at right angles both to OP and to E_2 and is equal to E_2 in magnitude.

Thus the part of the electromagnetic field which depends on the acceleration of the particle is specified by two vectors, the electric and magnetic intensities. These are mutually perpendicular and, in our units, are equal to one another in magnitude. They are both at right angles to the radius from the point P to the position O of the particle at the instant at which the state of its motion determined the field at P. The vectors may thus be said to be at right angles to the line of flight of the electromagnetic disturbance.

There is another very important difference between the part of the field which depends on the acceleration and that which does not. We saw that the electric and the magnetic intensity in the latter were both inversely proportional to the square of the distance from the moving charge. In the former both the intensities are inversely as the first power of this distance. So that at great distances from the moving charge the part of the field which depends on the acceleration will become very great compared with the part which does not depend upon it. The importance of this will be clearer after we have considered the distribution of the energy in the field.

Energy in the Field.

We know that the electromagnetic energy per unit volume of the field at any point is

$$\begin{aligned} U = T + W &= \tfrac{1}{2}\{H^2 + E^2\} \\ &= \tfrac{1}{2}\{\overline{H_1 + H_2}^2 + \overline{E_1 + E_2}^2\} \\ &= \tfrac{1}{2}\{H_1^2 + E_1^2\} + (H_1 H_2) + (E_1 E_2) + \tfrac{1}{2}\{H_2^2 + E_2^2\}, \end{aligned}$$

where $(H_1 H_2)$ and $(E_1 E_2)$ denote the scalar products of the vectors. The energy per unit volume may thus be split up into three parts,

$$U_1 = \tfrac{1}{2}\{H_1^2 + E_1^2\}, \qquad U_2 = (H_1 H_2) + (E_1 E_2),$$

and

$$U_3 = \tfrac{1}{2}\{H_2^2 + E_2^2\}.$$

Following Langevin* we shall consider these three parts of the energy separately.

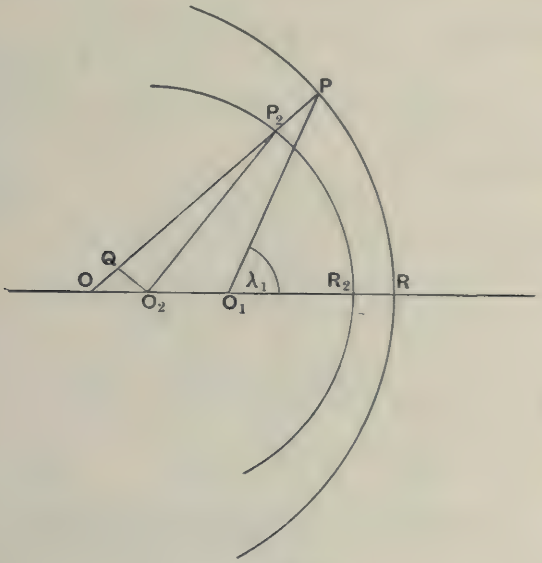


Fig. 34.

---

* *Journal de Physique*, vol. IV. p. 171 (1905).

We have seen that the field at the point P at the time t is determined by the state of the moving charge at the instant $t-\theta$, when it was at a point O distant r from P, where $OP=r=c\theta$. The same statement is true of the field at all points on a sphere through P with O as centre. The field may thus be regarded as spreading out from the moving charge in all directions with the uniform velocity c. If we consider another point P_2 on OP the field at P_2 will no longer be determined by the state of the moving charge when at O, because the disturbance initiated at O has already passed over P_2 and reached P. The field at P_2 at the instant t will be determined by the state of the particle when at some point O_2 at a time later than $t-\theta$, let us say $t-\theta+d\theta$. The locus of the points at which the field at the time t is determined by the state of the particle at the time $t-\theta+d\theta$ is the sphere P_2R_2 whose centre is O_2 and whose radius is

$$O_2P_2=c\,(\theta-d\theta).$$

At the time t the state of the field in the excentric shell bounded by the two excentric spheres whose radii are

$$OP=c\theta \quad\text{and}\quad O_2P_2=c\,(\theta-d\theta)$$

will be determined, at every point, by the state of the moving charge at some instant between the times $t-\theta$ and $t-\theta+d\theta$. We shall now consider the energy of each of the three types U_1, U_2 and U_3 which is present at the instant t inside the excentric shell.

The energy of the first type per unit volume is

$$\begin{aligned}
U_1&=\tfrac{1}{2}\{E_1^2+H_1^2\}=\tfrac{1}{2}E_1^2(1+\beta^2\sin^2\lambda_1)\\
&=\frac{e^2(1-\beta^2)^2}{32\pi^2}\,\frac{1+\beta^2\sin^2\lambda_1}{r_1^4(1-\beta^2\sin^2\lambda_1)^3}\\
&=\frac{e^2(1-\beta^2)^2}{32\pi^2}\,\frac{r_1^2(1+\beta^2\sin^2\lambda_1)}{r^6(1-\beta\cos\lambda)^6}\\
&=\frac{e^2(1-\beta^2)^2}{32\pi^2}\,\frac{r_1^2+\beta^2r^2\sin^2\lambda}{r^6(1-\beta\cos\lambda)^6}\\
&=\frac{e^2(1-\beta^2)^2}{32\pi^2r^4}\,\frac{1+2\beta^2-2\beta\cos\lambda-\beta^2\cos^2\lambda}{(1-\beta\cos\lambda)^6},
\end{aligned}$$

by making use of the various relations on pp. 248 and 249.

The energy of this type within the excentric shell is $\iiint U_1\,d\tau$

taken throughout the volume of the shell. Using polar coordinates r, λ, ϕ the annular element of volume is

$$2\pi r \sin\lambda \times r d\lambda \times PP_2.$$

Now $OP = c\theta$, $O_2P_2 = c(\theta - d\theta)$, and if Q is the foot of the perpendicular from O_2 on OP,

$$OQ = OO_2 \cos\lambda = Vd\theta \cos\lambda = c\beta \cos\lambda d\theta.$$

In the limit when $d\theta$ is made to vanish, $QP_2 = O_2P_2$. Thus

$$\begin{aligned} PP_2 &= c\theta - c(\theta - d\theta) - c\beta\cos\lambda d\theta \\ &= c d\theta (1 - \beta\cos\lambda). \end{aligned}$$

Hence

$$\begin{aligned} \iiint U_1 d\tau &= \frac{e^2(1-\beta^2)^2}{16\pi r^2} c d\theta \int_0^\pi \frac{1 + 2\beta^2 - 2\beta\cos\lambda - \beta^2\cos^2\lambda}{(1-\beta\cos\lambda)^5} \sin\lambda d\lambda \\ &= \frac{e^2(1-\beta^2)^2}{16\pi r^2} \frac{c d\theta}{\beta} \int_{1-\beta}^{1+\beta} \frac{2\beta^2 - 2 + 4x - x^2}{x^5} dx, \\ &\qquad\qquad (\text{where } x = 1 - \beta\cos\lambda), \\ &= \frac{e^2}{24\pi r^2} \frac{3+\beta^2}{1-\beta^2} c d\theta \quad \text{.........}(7). \end{aligned}$$

This is the value of the part of the energy, which does not depend upon the acceleration, which is found at the time t between the excentric spheres whose radii are $c\theta$ and $c(\theta - d\theta)$. It will be observed that for a constant value of $c d\theta$ this part of the energy varies inversely as the square of r. It is therefore negligible at a great distance and is practically all concentrated in the immediate neighbourhood of the moving charge.

In the case in which the charge has always been in motion with the uniform rectilinear velocity $V = \beta c$, all the energy is of this type. The formula we have just obtained enables us to calculate the total energy of this type which lies outside a small sphere of radius R described about the moving charge. The smallest of the excentric spheres will have the instantaneous position of the moving charge, that is to say its position at the time t, as centre.

All the other spheres, see Fig. 35, are external to this one and their centres behind its centre O_0, since we are supposing that the charge is moving with a velocity which is smaller than that of light. We may now, in imagination, displace all the spheres so that they have the point O_0 as centre. When this is done the volume of the shell between any two spheres is the same as when they were excentric. The total energy in the field is clearly the same as if the energy of each excentric shell was spread uniformly over the volume of the corresponding concentric shell. It is therefore

$$\frac{e^2}{24\pi}\frac{3+\beta^2}{1-\beta^2}\int_R^\infty \frac{c\,d\theta}{r^2} = \frac{e^2}{24\pi}\frac{3+\beta^2}{1-\beta^2}\int_R^\infty \frac{dr}{r^2}$$

$$= \frac{e^2}{24\pi R}\frac{3+\beta^2}{1-\beta^2} = \frac{e^2}{8\pi R}\left\{1+\frac{4\beta^2}{3(1-\beta^2)}\right\} \quad \text{...(8).}$$

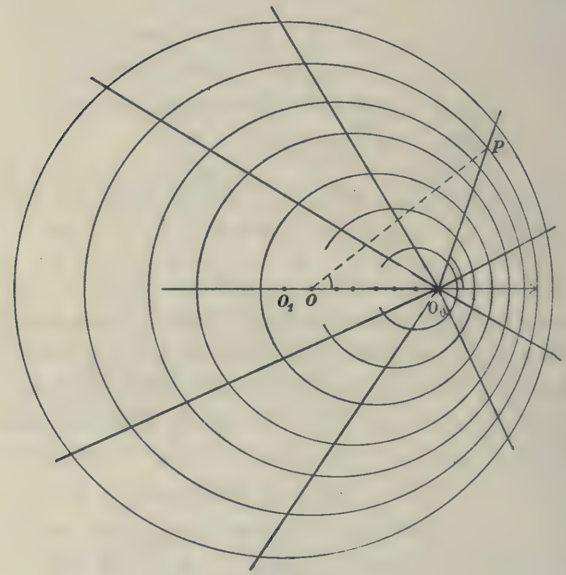


Fig. 35.

In our units the potential (electrostatic) energy outside the sphere of radius R is $e^2/8\pi R$. The remainder must therefore be the energy of the magnetic field of the moving charge. Since the electrostatic energy is the same whether the particle is at rest or in motion, the magnetic energy is the same thing as the kinetic energy of the moving charge. The kinetic energy is thus

$$\frac{e^2}{6\pi R}\frac{\beta^2}{1-\beta^2} \quad \text{..........................(9).}$$

This may be regarded as the part of the kinetic energy of the moving point charge which lies outside the sphere of radius R described about it as centre. In the case of an extended distribution of electric charge such as the spherical shell considered in the last chapter, it is necessary to calculate the combined effects of the superposed fields due to the different elements. The calculations are very complicated and can only lead to the results obtained previously, so that we shall not pursue the matter further in this direction.

It is important, however, to emphasize at this stage the two main features of the part of the energy of the field of the moving charge which does not involve the acceleration. In the first place the whole of this energy is located in the immediate neighbourhood of the moving charge and is carried along with it in its motion. In the second place its magnitude is a function of the geometrical distribution of the charge. Both are clearly established by the discussion immediately preceding and by the results of the last chapter. This part of the field has been called by Langevin the velocity wave of the moving charge.

We shall now consider the part of the energy which depends only on E_2 and H_2. The amount of this which is present between the two excentric spheres whose radii are $c\theta$ and $c(\theta - d\theta)$ may be calculated as follows. Since $E_2 = H_2$ the energy per unit volume is

$$\tfrac{1}{2}(E_2^2 + H_2^2) = E_2^2.$$

Now E_2 is the resultant of two components, one along O_1P and

$$= \frac{e\Gamma \cos \mu}{4\pi c^2 r^2 (1 - \beta \cos \lambda)^3}\, \overline{O_1P},$$

and the other parallel to $\overline{\Gamma}$ and

$$= -\frac{e}{4\pi c^2 r (1 - \beta \cos \lambda)^2}\, \overline{\Gamma}.$$

Using the polar coordinates r, λ and ϕ, take the line OO_1 as the polar axis and choose the plane $\phi = 0$ so that it contains OO_1 and a line through O parallel to the acceleration $\overline{\Gamma}$ (Fig. 36). To find the value of E_2^2 resolve E_2 into three components, (1) X parallel to OO_1, (2) Y perpendicular to OO_1 and parallel to the plane $\phi = 0$, and (3) Z perpendicular to OO_1 and also to the plane $\phi = 0$. Then since $OO_1 = \beta r$ we have

$$\left.\begin{aligned} X &= \frac{e\Gamma}{4\pi c^2 r(1-\beta\cos\lambda)^2}\left\{\frac{\cos\mu(\cos\lambda-\beta)}{1-\beta\cos\lambda} - \cos\epsilon\right\} \\ Y &= \frac{e\Gamma}{4\pi c^2 r(1-\beta\cos\lambda)^2}\left\{\frac{\cos\mu\sin\lambda\cos\phi}{1-\beta\cos\lambda} - \sin\epsilon\right\} \\ Z &= \frac{e\Gamma}{4\pi c^2 r(1-\beta\cos\lambda)^2}\,\frac{\cos\mu\sin\lambda\sin\phi}{1-\beta\cos\lambda} \end{aligned}\right\} \ldots(10),$$

where ϵ is the angle between $\overline{\Gamma}$ and OO_1. We have now to evaluate the integral

$$\iiint (X^2 + Y^2 + Z^2)\, d\tau',$$

where $$d\tau' = r^2(1-\beta\cos\lambda)\sin\lambda\, d\lambda\, d\phi\, c\, d\theta.$$

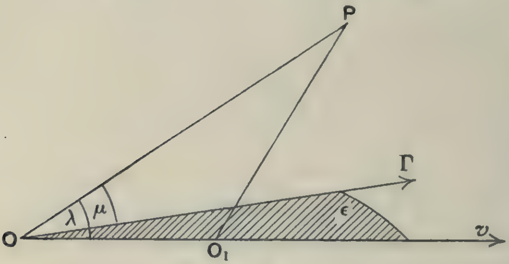


Fig. 36.

The limits of integration are: for λ, from 0 to π and for ϕ, from 0 to 2π.

Since $$\cos\mu = \cos\lambda\cos\epsilon + \sin\lambda\sin\epsilon\cos\phi,$$

and $$\int_0^{2\pi} d\phi = 2\pi, \quad \int_0^{2\pi}\cos\phi\, d\phi = 0,$$

and $$\int_0^{2\pi}\cos^2\phi\, d\phi = \pi,$$

after changing the other variable from λ to $x = 1 - \beta\cos\lambda$, we find, for the energy of this type between the excentric spheres, the value

$$dU_3 = \frac{e^2\Gamma^2}{6\pi c^3}\,\frac{1-\beta^2\sin^2\epsilon}{(1-\beta^2)^3}\, d\theta \;\ldots\ldots\ldots\ldots\ldots(11).$$

This result is extremely important. Since it does not involve r it shows that the energy between the spheres remains constant provided they always correspond to the same instants $t-\theta$ and $t-\theta+d\theta$. Thus as the spheres travel outwards with the velocity c they always enclose the same amount of this energy. *The total energy of this type which is emitted by the accelerated charged particle travels off, in undiminished amount, in all directions to infinity with the uniform velocity c.* This energy can no longer be regarded as belonging to the moving system. It is transferred to the surrounding space and is in fact the energy *radiated* by the particle.

Since dU_3 is proportional to Γ^2 we see that there is no radiation except when the moving charge is accelerated. There is no radiation from a particle in uniform motion. Moreover the radiation has the same value whether Γ is positive or negative. A retardation has the same effect as an acceleration. This is true not only for small velocities but also for velocities which are comparable with that of light.

Our results may readily be extended from a point charge to any finite geometrical distribution. By considering two spheres such that both r and $dr\,(=c\,d\theta)$ are large compared with the dimensions of the charged system, it is clear that the radiation is independent of the distribution of the charge within the system, provided the acceleration is the same for all parts of the system. This establishes an important difference between the radiant energy and the part of the energy which is independent of the acceleration. For we saw that the latter part depended very directly on the geometrical distribution of the electrification, being in fact a linear function of the electromagnetic mass of the system. On the other hand, *the energy radiated by a charged system does not involve the electromagnetic mass of the system.*

When the system contains charges of both signs, and also when different parts have different accelerations, the radiation will in general depend on the geometrical distribution; but, in any event, the influencing factors are not those which determine the electromagnetic mass. It is clear that so far as the radiation which escapes to a great distance is concerned each electron in any material system may be treated as a point charge. When the velocities are small the energy in the shell does not involve the angle ϵ and it may be shown, in a manner rather similar to the foregoing calculations, that the radiation is proportional to $(\Sigma e\Gamma)^2$, taken over all the electrons or charged particles in the system. It follows from this result that a material system which is electrically neutral will not emit radiation when it is uniformly accelerated, even though it is made up of electrons. This result is otherwise obvious since it is clear that any system which is arranged so as to have no external field will not radiate. The result that the radiation is proportional to $(\Sigma e\Gamma)^2$ enables us to see at a glance whether a given system will be an efficient radiator or not. For instance, we may take the case of two equal and opposite charges revolving in a circular orbit about the mid-point of the line joining their centres. Here both the charges and the accelerations have opposite signs; so that the sum of the products is additive and the system is a good radiator. In the case of two equal negative charges revolving in the same orbit about a positive charge at its centre the charges have the same sign, and their accelerations opposite signs. The value of $(\Sigma e\Gamma)^2$ is therefore zero and there is no radiation from this system. There will of course be radiation emitted if the negative charges get out of phase with each other. It is well to observe that the vanishing of $(\Sigma e\Gamma)^2$ will be the condition for no radiation only provided that the summation is taken over a sufficiently small element of volume. The result that the rate of radiation of energy from a point charge is equal to

$$\frac{e^2\Gamma^2}{6\pi c^2} \qquad \text{.............................(12)}$$

for small velocities was first given by Larmor*.

Some of the properties of the radiation waves will be considered in later chapters.

The Wave of Reorganization.

The part dU_2, of the energy in the shell, which involves the geometrical products of E_1 and E_2, and of H_1 and H_2 may be obtained in a manner similar to that used in calculating the value of dU_3. Taking the same Cartesian axes as before, we find the values of the components X_1, Y_1, Z_1 of E_1 and P_1, Q_1, R_1 of H_1 from the expressions on p. 248. We have already found X_2, Y_2, Z_2

---

* *Phil. Mag.* V. vol. XLIV. p. 503 (1897). Cf. also *Aether and Matter*, p. 227.

the components of E_2, and P_2, Q_2, R_2 may readily be obtained since we know that E_2 and H_2 are equal to each other and are also perpendicular both to each other and to OP. The value of this part of the energy per unit volume is

$$[X_1X_2 + Y_1Y_2 + Z_1Z_2 + P_1P_2 + Q_1Q_2 + R_1R_2].$$

Introducing the element of volume in polar coordinates we integrate this throughout the region between the two shells and find

$$dU_2 = \frac{1}{(1-\beta^2)^2}\frac{e^2}{3\pi r}\frac{\beta\Gamma}{c}\cos\epsilon\, d\theta \quad \text{............(13).}$$

This energy is a maximum when $\epsilon = 0$ and is zero when $\epsilon = \pi/2$. Since $\Gamma d\theta$ is the change of velocity in the interval $d\theta$ we see that $\Gamma d\theta \cos\epsilon$ is equal to the resolved part of the change of velocity in the direction of motion. Calling this $c\delta\beta$ we have

$$dU_2 = \frac{e^2}{3\pi r(1-\beta^2)^2}\beta d\beta \quad \text{..................(14).}$$

This part of the energy contains r in the denominator, so that it falls to zero as the two spheres proceed to infinity.

There is a rather simple and important relation between the energy dU_2 and the energy in the field when the charge is in uniform motion. In considering this, in order to fix our ideas, we shall suppose that the motion is uniform except during the infinitesimal interval between $t-\theta$ and $t-\theta+d\theta$. Before $t-\theta$ the charge moves in a straight line with uniform velocity v_1 and after $t-\theta+d\theta$ it again moves uniformly in a straight line but with a different velocity v_2. Consider the energy in the field outside of the two excentric spheres. Up to the instant t this is that which is proper to the case of a charge moving with the uniform velocity v_1. At the instant t this begins to be altered. A new field begins to be established, namely that which is proper to a particle moving with the uniform velocity v_2. The new field is evidently left behind it as the excentric shell travels outwards with the velocity c. The relation referred to is that the energy dU_2 is just what is required to change the field *outside* of the excentric sphere from the state corresponding to the uniform velocity v_1 to the state corresponding to v_2. For this reason the energy dU_2 has, very appropriately, been called by Langevin the wave of reorganization.

The proof of this assertion is a very simple matter. We have seen (p. 254) that for a point charge moving with uniform velocity $v_1 = c\beta$ the energy $U_1(\beta)$ of the field outside a sphere of radius r about the position of the particle at $t-\theta$ is

$$U_1(\beta) = \frac{e^2}{8\pi r}\left(1 + \frac{4}{3}\frac{\beta^2}{1-\beta^2}\right).$$

If the speed instead of being v_1 were $v_2 = c(\beta + \delta\beta)$, let us say, the value of U_1 would be

$$\begin{aligned} U_1(\beta+\delta\beta) &= \frac{e^2}{8\pi r}\left(1 + \frac{4}{3}\frac{(\beta+\delta\beta)^2}{1-(\beta+\delta\beta)^2}\right) \\ &= \frac{e^2}{8\pi r}\left(1 + \frac{4}{3}\frac{\beta^2}{1-\beta^2} + \frac{8}{3}\frac{\beta\delta\beta}{(1-\beta^2)^2}\right), \end{aligned}$$

neglecting higher powers of $\delta\beta$ than the first. The difference $U_1(\beta+\delta\beta) - U_1(\beta)$ is equal to the value of dU_2 given by (14). The extension to a finite change of velocity spread over a finite interval of time is obvious.

Acceleration and Force.

Let us consider the work done by an external force on an electric charge. The law of conservation of energy requires that this should be equal to the gain in the energy of the system, and we have seen that the whole energy of the system can be expressed as energy of the electromagnetic field. The increment of energy consists in fact of two parts: the first dU_2 which is required to change the state of the electric field to the condition permanently appropriate to the new state of motion, and the second dU_3 which represents the energy transformed into radiation. If Xe is the force acting on the charge we have therefore

$$\begin{aligned} Xe\delta x &= \delta U_2 + \delta U_3 \\ &= \frac{1}{(1-\beta^2)^2}\frac{e^2}{3\pi r}\frac{\beta}{c}\Gamma\cos\epsilon\, d\theta + \frac{1-\beta^2\sin^2\epsilon}{(1-\beta^2)^3}\frac{e^2}{6\pi c^3}\Gamma^2\, d\theta \\ &= \frac{e^2}{3\pi c^2 r(1-\beta^2)^2}\Gamma\,\delta x + \frac{1-\beta^2\sin^2\epsilon}{(1-\beta^2)^3}\frac{e^2}{6\pi c^2}\frac{1}{c\beta\cos\epsilon}\Gamma^2\,\delta x \quad \ldots(15), \end{aligned}$$

since $\delta x = c\beta\cos\epsilon\, d\theta$. The right-hand side of this equation would represent the whole work of the force in the hypothetical case of an electron whose properties were those arising from that part of the field due to a point charge e, which lay outside a sphere of radius r about the charge. It is very unlikely that an electron possesses this constitution, and where the electrification has an extended geometrical configuration the equation has to be modified. We have seen that the second term on the right-hand which arises from the energy radiated is not affected by the geometrical arrangement of the charge. With the first term it is different. The new value of this may be found by considering the difference in the energy in the steady fields corresponding to the velocities before and after the action of the force. Proceeding in this way we find, in the case of any distribution possessing spherical symmetry, that r is replaced by the radius of the sphere multiplied by a numerical factor. The precise value of the factor depends on the configuration of the distribution, whether superficial or throughout the volume of the sphere and, if so, whether uniform or not. Thus the form of the equation is unaltered even when the electric charge is not concentrated at a point.

Our equation differs from the corresponding Newtonian Equation $F = m\Gamma$ in two important respects. In the first place we have the additional term depending upon Γ^2 and in the second place m is never constant, although it is approximately so when β is very small. This aspect of the question has already been considered. When the acceleration is small the second term containing Γ^2 will be relatively unimportant, so that the Newtonian law is an approximation which is true for small accelerations and small velocities.

For the *form* of the Newtonian law to be preserved it is necessary that the second term should be negligible compared with the first. Thus if a is the radius of the electron it is necessary that

$$\frac{\delta U_3}{\delta U_2} = \frac{(1-\beta^2 \sin^2 \epsilon)\,\Gamma a}{2\,(1-\beta^2)\,c^2\beta \cos \epsilon} \quad \text{...............(16)}$$

should be small compared with unity. Since $1-\beta^2$ may in all practical cases be taken to be of the order unity, the order of magnitude of this fraction is

$$\frac{\Gamma a}{2c^2\beta \cos \epsilon} = \frac{1}{2c}\,\Gamma\,\frac{a}{c\beta \cos \epsilon} = \frac{1}{2c}\,\Gamma\,\frac{a}{v \cos \epsilon} = \frac{\Gamma \tau}{2c} \text{......(17)},$$

where τ is the time necessary for the charged sphere to move in the direction of the acceleration through a distance equal to its own radius. It is therefore necessary that the change in the velocity of the body during this time τ should be small compared with the velocity of radiation.

If this condition is satisfied the motion will be the same as that calculated on the supposition that radiation does not occur. Abraham[*] has suggested that the term *quasi-stationary* be applied to states of motion which satisfy this condition. Since the calculation of the motion when the state is not quasi-stationary is extremely complicated, it is important to inquire whether such states are likely to occur in nature or not. There is one case which is of very frequent occurrence in which the condition is violated, and that is the case in which the velocity of the particle and β are zero. This state, however, only lasts for an insignificant interval of time in the case of an electron. We know from the value of the electromagnetic mass of these particles that a is about 10^{-13} cm. If the particle has moved from rest for a time t the velocity $v = \Gamma t$, and since $c = 3 \times 10^{10}$ cms. per sec. the fraction $\frac{\Gamma a}{2cv \cos \epsilon}$ will be comparable with $10^{-13}/(6 \times 10^{10} \times t)$. This will clearly be negligible compared with unity for any measurable interval of time. Even after 10^{-20} sec. it will have fallen almost to 10^{-4}. Thus even in the case of a charged particle starting from rest it seems unlikely that serious error will arise if the motion is treated as though it were quasi-stationary.

Another case in which one might expect the quasi-stationary condition to be departed from arises when the acceleration is very great. Probably the greatest acceleration with which we are familiar is that which occurs during the impact of a β-ray particle on an atom. There is some evidence that in favourable cases a β-ray particle may be completely stopped within a distance d equal to the diameter of an atom, let us say 10^{-8} cm. Assuming uniform acceleration as an illustration, the equation of energy in this case is

$$\tfrac{1}{2}mv^2 = m\Gamma d.$$

Thus
$$\Gamma = v^2/(2 \times 10^{-8}),$$

and
$$\frac{\Gamma a}{2cv} = \tfrac{1}{4}\,(10^{-5})\,\frac{v}{c} = \tfrac{1}{4} \times 10^{-5},$$

---

[*] *Ann. der Phys. loc. cit.*

when v has its maximum possible value c. Even in this case the ratio has a value which is quite small, so that it does not seem likely that the assumption of quasi-stationary motion will lead to serious error in practical cases.

The Reaction of the Radiation.

Although the part of the force which is neglected by the assumption of quasi-stationary motion is practically always small compared with the remaining part, there are cases in which it might exert important effects through its persistency. For example in the case of periodic motions, if the reaction of the radiation always tended to stop the motion, it would in time exert an appreciable effect. For this reason it is desirable to calculate the reaction on an accelerated electron without assuming, as we did at the beginning of this chapter, that the increase in the velocity during the time required by the electron to move over its own radius is small compared with the velocity of radiation. Owing to the difficulties which arise in a more general treatment we shall content ourselves with the case in which the velocity, though variable, is rectilinear, and in which the squares and higher powers of the ratios of the velocity of the electron, and the derivatives of the velocity, to the velocity of light may be neglected.

Let us seek the force acting on an element of volume $d\tau'$ of the moving charged sphere at the point P whose coordinates are x', y', z' at the instant t. Let the variable velocity u of the sphere be parallel to the axis of x. Let Q (x, y, z) be the position of some other point of the sphere at the time t. The part of the potentials at P which arises from the element $d\tau = dx\,dy\,dz$ which is at the point Q at the time t will be determined, not by the instantaneous state and position Q of the element of volume, but by its state and position $Q_\theta = x_\theta$, y, z at a time $t - \theta$, where

$$PQ_\theta = \{(x_\theta - x')^2 + (y - y')^2 + (z - z')^2\}^{\frac{1}{2}} = c\theta.$$

We also have

$$PQ = \{(x - x')^2 + (y - y')^2 + (z - z')^2\}^{\frac{1}{2}} = r,$$

and

$$x_\theta = x - \frac{\partial x}{\partial t}\theta + \frac{\partial^2 x}{\partial t^2}\frac{\theta^2}{2} - \frac{\partial^3 x}{\partial t^3}\frac{\theta^3}{6} + \dots$$

$$= x - u\theta + \tfrac{1}{2}\dot{u}\theta^2 - \tfrac{1}{6}\ddot{u}\theta^3 + \dots .$$

Since $c\theta$ and r differ only by terms involving u and its derivatives as a factor we can put $r = c\theta$ in the equation for x_θ.

Thus
$$x_\theta = x - \frac{u}{c}r + \frac{\dot{u}}{2c^2}r^2 - \frac{\ddot{u}}{6c^3}r^3 + \dots .$$

Substituting this value of x_θ in PQ_θ, we obtain

$$PQ_\theta^2 = c^2\theta^2 = (y-y')^2 + (z-z')^2 + \left\{\frac{u}{c}r - \frac{\dot{u}}{2c^2}r^2 + \frac{\ddot{u}}{6c^3}r^3 - \dots - (x-x')\right\}^2$$
$$= r^2 - 2(x-x')\left(\frac{u}{c}r - \frac{\dot{u}}{2c^2}r^2 + \frac{\ddot{u}}{6c^3}r^3 - \dots\right)$$

+ terms involving squares of small quantities.

Thus
$$c\theta = r\left\{1 - 2\frac{x-x'}{r}\left(\frac{u}{c} - \frac{\dot{u}}{2c^2}r + \frac{\ddot{u}}{6c^3}r^2 - \dots\right)\right\}^{\frac{1}{2}}$$
$$= r\left\{1 - \frac{x-x'}{r}\left(\frac{u}{c} - \frac{\dot{u}}{2c^2}r + \frac{\ddot{u}}{6c^3}r^2 - \dots\right)\right\}$$

+ higher terms.

We also have
$$\frac{\partial x_\theta}{\partial x} = 1 - \frac{u}{c}\frac{\partial r}{\partial x} + \frac{\dot{u}}{c^2}r\frac{\partial r}{\partial x} - \frac{\ddot{u}}{2c^3}r^2\frac{\partial r}{\partial x} + \dots .$$

But
$$\frac{\partial r}{\partial x} = \frac{x-x'}{r},$$

so that
$$dx_\theta = \left(1 - \frac{u}{c}\frac{x-x'}{r} + \frac{\dot{u}}{c^2}(x-x') - \frac{\ddot{u}}{2c^3}r(x-x') + \dots\right)dx.$$

Thus to every element of volume $d\tau\,(= dx\,dy\,dz)$ of the charge at the time t there will be the corresponding element

$$d\tau_\theta\,(= dx_\theta\,dy\,dz)$$

at the time θ seconds previously, where

$$d\tau_\theta = \left(1 - \frac{u}{c}\frac{x-x'}{r} + \frac{\dot{u}}{c^2}(x-x') - \frac{\ddot{u}}{2c^3}r(x-x') + \dots\right)d\tau.$$

The velocity of this charge at the time $t - \theta$ is

$$u - \dot{u}\theta + \tfrac{1}{2}\ddot{u}\theta^2 - \tfrac{1}{6}\ddot{u}\theta^3 + \dots$$
$$= u - \dot{u}\frac{r}{c} + \tfrac{1}{2}\ddot{u}\frac{r^2}{c^2} - \tfrac{1}{6}\ddot{u}\frac{r^3}{c^3} + \dots .$$

In integrating for the scalar and vector potentials we must replace $\dfrac{d\tau}{r}$ by $\dfrac{d\tau_\theta}{c\theta}$. On substituting the values already obtained, we find to the first order of small quantities that

$$\frac{d\tau_\theta}{c\theta} = \frac{d\tau}{r}\left\{1 + \frac{\dot{u}}{2c^2}(x - x') - \frac{\ddot{u}}{3c^3}(x - x')\,r + \ldots\right\}.$$

The scalar potential

$$\phi = \frac{1}{4\pi}\iiint\left(\rho\,\frac{d\tau}{r}\right)'$$

is evidently

$$= \frac{1}{4\pi}\iiint\frac{\rho}{r}\left\{1 + \frac{\dot{u}}{2c^2}(x - x') - \frac{\ddot{u}}{3c^3}(x - x')\,r + \ldots\right\}d\tau.$$

The vector potential

$$U_x = \frac{1}{4\pi c}\iiint\left(\rho\,\frac{u\,d\tau}{r}\right)'$$

contains u as a factor, so that the replacement of $\dfrac{d\tau}{r}$ by $\dfrac{d\tau_\theta}{c\theta}$ will only introduce small quantities of the second order and may be omitted. It is, however, necessary to substitute in U_x the value of u at the instant $t - \theta$. Thus

$$U_x = \frac{1}{4\pi c}\iiint\frac{\rho}{r}\left\{u - \frac{\dot{u}}{c}\,r + \tfrac{1}{2}\,\frac{\ddot{u}}{c^2}\,r^2 - \ldots\right\}d\tau.$$

$U_y = U_z = 0$, since the resultant velocity is parallel to the axis of x.

The x component of the force on a unit charge at P is

$$F_x = -\frac{\partial\phi}{\partial x'} - \frac{1}{c}\,\frac{\partial}{\partial t}(U_x) + \frac{1}{c}\,[uH].$$

The magnetic force $H\,(=\mathrm{rot}\,U)$ is derived by differentiating the vector potential with respect to the coordinates. It thus contains u as a factor, so that the lowest term in $[uH]$ will involve squares of small quantities. This term may therefore be neglected. Evidently

$$\frac{\partial\phi}{\partial x'} = -\frac{1}{4\pi}\iiint\frac{\rho}{r^2}\,\frac{\partial r}{\partial x'}\,d\tau + \frac{\dot{u}}{8\pi c^2}\iiint\rho\left\{-\frac{x - x'}{r^2}\,\frac{\partial r}{\partial x'} - \frac{1}{r}\right\}d\tau$$
$$+ \frac{\ddot{u}}{12\pi c^3}\iiint\rho\,d\tau + \ldots$$

$$= \frac{1}{4\pi}\iiint\rho\,\frac{x - x'}{r^3}\,d\tau + \frac{\dot{u}}{8\pi c^2}\iiint\rho\left\{\frac{(x - x')^2}{r^3} - \frac{1}{r}\right\}d\tau$$
$$+ \frac{\ddot{u}}{12\pi c^3}\iiint\rho\,d\tau + \ldots,$$

and

$$\frac{\partial U_x}{\partial t}=\frac{\dot{u}}{4\pi c}\iiint\frac{\rho}{r}\,d\tau-\frac{\ddot{u}}{4\pi c^2}\iiint\rho\,d\tau+\frac{\dddot{u}}{8\pi c^3}\iiint\rho r\,d\tau+\dots.$$

Thus

$$F_x=-\frac{1}{4\pi}\iiint\rho\,\frac{x-x'}{r^3}\,d\tau-\frac{\dot{u}}{8\pi c^2}\iiint\rho\left(\frac{1}{r}+\frac{(x-x')^2}{r^3}\right)d\tau$$
$$+\frac{\ddot{u}}{6\pi c^3}\iiint\rho\,d\tau+\dots.$$

This is the force on a unit charge at P arising from the whole sphere. The force on an element of volume $d\tau'$ at P is therefore

$$\delta F_x'=-\rho' d\tau'\left\{\frac{1}{4\pi}\iiint\rho\,\frac{x-x'}{r^3}\,d\tau+\frac{\dot{u}}{8\pi c^2}\iiint\rho\left(\frac{1}{r}+\frac{(x-x')^2}{r^3}\right)d\tau\right.$$
$$\left.-\frac{\ddot{u}}{6\pi c^3}\,e+\dots\right\},$$

since $e=\iiint\rho\,d\tau$. The total force on the whole moving charge due to itself is

$$F_x'=-\frac{1}{4\pi}\iiint\rho' d\tau'\iiint\rho\,\frac{x-x'}{r^3}\,d\tau$$
$$-\frac{1}{8\pi c^2}\iiint\dot{u}\rho' d\tau'\iiint\rho\left(\frac{1}{r}+\frac{(x-x')^2}{r^3}\right)d\tau+\frac{e}{6\pi c^3}\iiint\rho' d\tau'\ddot{u}+\dots.$$

The value of the first term when integrated over the whole sphere is zero, by symmetry. Transforming to polar coordinates the second term reduces, on integration, to $-\dfrac{e^2\dot{u}}{6\pi ac^2}$ in the case of a uniformly charged spherical shell of radius a and total charge e, whilst the third term is $\dfrac{e^2\ddot{u}}{6\pi c^3}$. Thus

$$F_x'=-\frac{e^2\dot{u}}{6\pi ac^2}+\frac{e^2\ddot{u}}{6\pi c^3}\quad\dots\dots\dots\dots\dots\dots(18).$$

The term in $\dot{u}$ appeared in the case of quasi-stationary motion. The term in $\ddot{u}$ is new and is therefore due to the fact that the quasi-stationary condition is not satisfied. The fact that it is independent of the geometrical distribution of the electrification at once suggests that it is due to the reaction of the radiation on the moving charge, since we have seen that the radiation itself possesses a similar property. This is made clearer if we consider the work done by the force F_x' during a displacement $u\,dt$.

This is $$F_x' u\,dt = -\frac{e^2}{6\pi ac^2}\,u\dot{u}\,dt + \frac{e^2}{6\pi c^3}\,u\ddot{u}\,dt$$
$$= -\frac{e^2}{6\pi ac^2}\frac{d}{dt}(\tfrac{1}{2}u^2)\,dt + \frac{e^2}{6\pi c^3}\left\{\frac{d}{dt}(u\dot{u}) - \dot{u}^2\right\}dt.$$

The term $-\frac{e^2}{6\pi ac^2}\frac{d}{dt}(\frac{1}{2}u^2)\,dt$ represents the diminution of the electromagnetic energy $(\frac{1}{2}mv^2)$ of the steady part of the field of the moving charged body. The term $-\frac{e^2}{6\pi c^3}\dot{u}^2 dt$ is equal to the energy radiated.

The work done by the second term in F_x' in a finite time, say from t_1 to t_2, is

$$\frac{e^2}{6\pi c^3}\left\{\Big| u\dot{u}\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2}\dot{u}^2 dt\right\}.$$

The integrated part vanishes if either u or $\dot{u}$ is zero at the limits. In the case of periodic motions this term will vanish when the limits are the recurrent zero values of u and $\dot{u}$. Even if these limits are not chosen its value over a long period of time will be small, being comparable with the maximum value within a single period. The value of $\dot{u}^2$, however, is always positive, when not zero, so that the integral $\int_{t_1}^{t_2}\dot{u}^2 dt$ will increase indefinitely as t_2 increases. It will evidently tend continuously to diminish the kinetic energy of the vibrating electrified particle, and thus to stop its motion. It is, in fact, very similar in effect to the action of a frictional force on a dynamical system.

Planck* has suggested that this force may account for the frictional term which it is necessary to introduce in order to account for absorption in optics. Whilst there is no doubt that this force must be operative it is, I think, far too small to account for any of the observed effects. The kind of action in which the frictional term is smallest is found in the case of substances which give rise to the residual rays. We have seen in Chapter VIII that even in these cases the coefficient which enters into the resistance term is of the order 10^{12} in the units employed, whereas the term $\frac{e^2\ddot{u}}{6\pi c^3}$ would lead to only about 10^{-7} of this amount. The main part of what appears to be a damping force in optical radiators must evidently be sought elsewhere.

---

* *Vorlesungen über die Theorie der Wärmestrahlung*, p. 109. Leipzig (1906).

CHAPTER XIII

THE AETHER

In our discussion of electromagnetic action so far we have always, explicitly or implicitly, considered the medium, in which the actions take place, to be at rest. It is true that in Chapters XI and XII we have considered in detail the effects produced by moving charges, and we have seen that changes in the state of the electromagnetic field are propagated through the surrounding medium with finite velocity. Without having specified the matter very definitely we have implicitly assumed that the bodies, whose motional effects were being investigated, were moving relatively to a fixed system. The fixed system embraces the observer and his instruments, and we have treated the question as though the medium through which the electromagnetic effects are propagated was rigidly attached to the observing system. This assumption is clearly, however, an arbitrary one; so far as anything which has yet been brought forward is concerned, we might just as well have considered the medium to be moving along with the moving system. It is necessary, then, to consider what effects we should expect to arise from the motion of the medium, in order to decide which, if either, of the foregoing alternatives is true. It may even happen that it is impossible to form a consistent scheme of electromagnetic phenomena, without discarding the idea of a medium altogether.

The main question at issue relates to the hypothetical non-material medium—the aether—which has long been supposed to be the seat of optical and electrodynamic actions in space. Although the question of the effects which are peculiar to moving material media is intimately connected with this, nevertheless the two questions are essentially different. If the first question is settled, we shall see (p. 285) that we have already accumulated all the material which is necessary to determine the specific effects which are caused by the motion of material media.

It appears to the writer to be impossible to acquire a true perspective of the matter at issue without considering it in its historical development. We shall therefore take it up from that point of view, at the risk perhaps of lengthening the treatment. Naturally, optical effects will be treated as a particular case of electrodynamic actions.

The Aberration of Light.

The study of this subject arose out of a discovery by Bradley in 1728, made during an investigation whose object was to detect annual parallax in certain fixed stars near the zenith. Such a parallax was found. It was not, however, directed towards the sun as it would have been if it were ordinary stellar parallax, but it was in a direction perpendicular to this in the plane of the earth's orbit. The magnitude of the "aberration" was found to be proportional to the sine of the inclination of the star, but was constant for stars of equal inclination. The results were found to be capable of complete explanation on the view that the light was propagated in space with a finite velocity in a direction which was fixed relatively to the star and which was uninfluenced by the earth's motion.

The problem is one of relative motion and can be made quite clear by considering an analogous material case. Suppose an observer to be in an open carriage which is moving with uniform velocity in a horizontal plane. He wishes to determine the direction of motion of the drops which fall into the carriage in a rain storm. To do this he is provided with a long tube which can rotate about a horizontal axis perpendicular to the length of the tube, and the inclination can be read off on a circle in a vertical plane. The direction is then determined by adjusting the tube so that the drops fall through without reaching the sides. The direction of the rain as thus determined will evidently depend on the direction of the rain relative to axes fixed in the ground, on its velocity and on that of the carriage relative to the same axes. Let us consider the comparatively simple case when the direction of the rain is in a vertical plane containing the direction of motion of the carriage.

Let OP be the direction of the tube, i.e. the apparent direction in which the rain reaches the carriage, and let the angle $OPP' = \alpha$. Then if $OP' : PP' = V : v$, where V is the velocity of the rain relative to the axes fixed in the earth, and v is the velocity of the carriage referred to the same axes, then OP' will represent the position of the path of the rain relative to axes fixed in the earth. For P' is then determined by the condition that the time required by the rain to move from O to P' is equal to that required by P to reach P'. Thus

$$\frac{v}{V} = \frac{PP'}{OP'} = \frac{\sin POP'}{\sin \alpha}.$$

The angle between the true and apparent directions is called the angle of aberration; denoting it by β we have

$$\beta = \angle POP' = \sin^{-1}\left(\frac{v}{V}\sin\alpha\right).$$

The analogy between this illustration and the optical case is complete. OP has to be replaced by the direction of the telescope and OP' by that of the light passing through it—relative to axes fixed in space. V becomes c the velocity of light in free space, and v is the velocity of the earth in its orbit. Thus

$$\sin\beta = \frac{v}{V}\sin\alpha,$$

and is therefore proportional to the sine of the apparent inclination of the star, as Bradley found. v/V is often referred to as the aberration constant.

If we make use of the hypothesis of the aether the most obvious interpretation of Bradley's observations is that the aether is at rest in space and is entirely uninfluenced by the motion of the earth through it. On this view, as was first pointed out by Boscovich, the angle of aberration will depend on the velocity of light in the medium in which the aberration takes place. If V' is the velocity of light in this medium the angle of aberration will be given by

$$\sin\beta = \frac{v}{V'}\sin\alpha.$$

It should therefore vary directly as the index of refraction of the medium. An experiment to test this conclusion was carried out long afterwards by Airy who used a telescope filled with water. He was, however, surprised to find that under these circumstances the constant of aberration had the same value as in an ordinary telescope.

An experiment based on somewhat similar reasoning had occurred independently to Arago, who argued that since the deviation of a ray of light produced by a dense prism depends on the ratio of the velocity of light in the prism and in the surrounding medium, its magnitude ought to be different according to whether the passage of the light through the prism is helped or retarded by the motion of the earth. With this experiment it is evidently unnecessary to use light from extra-terrestrial sources. Arago found that no effect due to the earth's motion could be detected, although the expected effect was comparatively large.

An explanation of the experimental results of both Airy and Arago was given in a general way by Fresnel, who suggested that the aether was carried along by moving material media in such a way as to compensate exactly for the difference between the velocities of the light in the medium and in vacuo. It is necessary that the aether should be carried along with a velocity which is only a fraction of that of the medium, for if it were carried with the same velocity there would be no aberration at all. Fresnel's suggestion was worked out more completely, later, by Maxwell and Stokes. We shall now calculate with what fraction of the velocity of the transparent medium it is necessary that the aether should be carried along in order to give the result required by Arago's experiment that refraction is independent of the earth's motion.

The path of a ray of light in any medium is determined by the fact that the time required to pass from any one point of the path to any other has a stationary value. This extension of Fermat's Principle of Least Time follows on the undulatory theory of light from the fact that if A and B are any two points in a ray the disturbance arising from points in the wave front in the immediate neighbourhood of A must all reach B in the same phase. This condition will evidently be satisfied when the time from A to B has either a stationary or a maximum value as well as when the value is a minimum. Thus if ds is an element of the path of the ray and V' the velocity at any point, then

$$\delta\int\frac{ds}{V'}=0\text{.............................(1).}$$

Consider the case of any optical system, including the observer, which moves through space with the velocity whose components are u, v, w. Let l, m, n be the direction cosines of the ray at any point. Suppose that in any refracting medium the velocity of the light passing through it is increased by the amount θ multiplied by the velocity of this medium through the aether. If V is the standard value of the velocity of light in this medium when the system is at rest, the velocity relative to axes fixed in space for the moving system will be

$$V+\theta\,(lu+mv+nw).$$

But the observer is moving relative to the fixed axes with a velocity whose components are u, v, w, so that the resolved part of his velocity along the direction of the ray is $lu+mv+nw$ and the velocity of the ray relative to him is

$$V-(1-\theta)(lu+mv+nw).$$

The equation of the relative ray paths is therefore

$$\delta\int\frac{ds}{V-(1-\theta)(lu+mv+nw)}=0,$$

or to the first order of $(u, v, w)/V$

$$\delta\int\frac{ds}{V}+\delta\int\frac{1-\theta}{V^2}(u\,dx+v\,dy+w\,dz)=0.$$

To be in agreement with Arago's experimental result it is necessary that the relative paths to this order should be independent of u, v, w. This will be the case if the quantity under the second integral sign is a perfect differential; since the value of the integral will then depend only on the terminal points which are not varied. If m is the refractive index of the medium and c is the velocity of light in free aether $m^2=c^2/V^2$. It is therefore necessary and sufficient that

$$\frac{m^2(1-\theta)}{c^2}(u\,dx+v\,dy+w\,dz)$$

be a perfect differential. Since the relation between m and u, v, w

is arbitrary and $udx + vdy + wdz$ satisfies this condition as a rule, it follows that

$$\frac{m^2(1-\theta)}{c^2} = \text{a constant } A,$$

so that $$\theta = 1 - Ac^2/m^2.$$

It is clear that θ must vanish for empty space for which $m = 1$, so that $A = 1/c^2$ and $\theta = 1 - 1/m^2$.

It follows therefore that there will be no effect, due to the motion of a refracting medium through space, of the order of magnitude of the first power of the ratio of the velocity of the medium to that of light; provided the velocity of light through the medium is increased by $(m^2-1)/m^2$ times the velocity of the medium. If we regard the optical disturbance as being propagated in an aether capable of flowing we may say that the aether is carried along by moving matter with $(m^2-1)/m^2$ times the velocity of the latter. This was Fresnel's interpretation of Arago's result.

It can also be shown that the hypothesis that the velocity of light in the moving medium is increased by $1-1/m^2$ of the velocity of the medium in space is sufficient to account for the fact that the aberration constant is independent of the refractive index of the medium filling the telescope which is used to measure it. For the discussion of the problem which is here involved the reader may be referred to Campbell, *Modern Electrical Theory*, First Edition, p. 293.

Fizeau's Experiment.

The conclusion that the velocity of light in a moving refracting medium is increased by $1-1/m^2$ of the velocity of the medium was put to the test of direct experiment by Fizeau. The apparatus he used for the purpose was similar to that shown in Fig. 37. Two parallel tubes $c\ c'$ were set up so that a stream of water could be made to run continuously through them in the directions indicated. A ray of light f was divided by the thinly silvered mirror a. The reflected portion travelled along the path $f\ a\ b\ c\ e\ c'\ b'\ a\ g$. The transmitted portion on the other hand followed the path $f\ a\ b'\ c'\ e\ c\ b\ a\ g$. The two beams thus combined in the direction of g and gave rise to interference bands which were observed with a telescope. It will be noticed that when the water is flowing one of the beams of light is always in the direction of flow of the water and the other against it. The position of the bands is first observed with the water at rest. The stream of water is then turned on and the displacement of the bands, caused by the resulting difference of velocity of the two beams, measured. Fizeau found that the shift of the bands thus caused was in complete accordance with Fresnel's hypothesis. The experiment has since been repeated with improved apparatus by Michelson and Morley, whose results were also in complete accordance with the view that the velocity of light in a moving material medium is increased by $1 - 1/m^2$ of the velocity of the medium.

The foregoing experimental results led to two rival views as to the relation between the motion of matter and that of the aether in its neighbourhood. The first view, which was championed by Fresnel, held that the aether outside material bodies was at rest, and matter moved through it without setting it in motion. Thus the aether was supposed to be able to flow freely through matter like water through a sieve. We have seen, however, that it is necessary to suppose that the relative velocity of moving refracting matter and the aether in its interstices depends on the refractive index as well as on the velocity of the matter relative to the stagnant aether. This was in agreement with Fresnel's optical ideas, according to which the density of the aether in material bodies depended on the refractive index. The relative velocity would therefore have to be different in order to preserve continuity of the medium. On this view aberration is the simplest possible problem in relative motion, and it is clear from

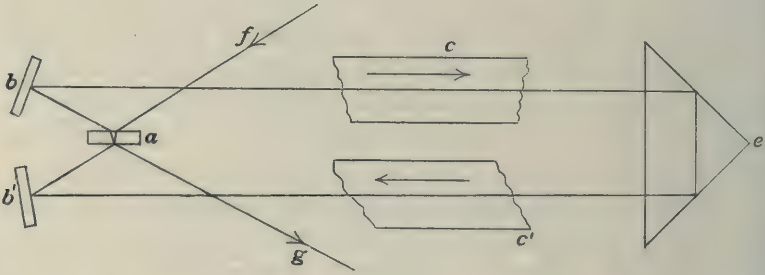


Fig. 37.

the foregoing discussion that Fresnel's views are in harmony with the experiments of Arago and Airy.

Stokes was unwilling to accept the view that matter could pass freely through the aether without setting it in motion, and he therefore undertook to investigate whether it would not be possible for the aether in its immediate neighbourhood to be carried along by the earth and still give the well-known results for terrestrial aberration. The problem is to find what distribution of velocity may be imparted to the aether so as to leave the paths of the rays in space unchanged. If c is the velocity of light in the stagnant aether in a direction whose cosines relative to fixed axes in space are l, m, n and the components of the supposed velocity of the aether are u, v, w at any point, then the velocity of the ray in space at that point will be $c + lu + mv + nw$. Applying the principle of stationary time the equations of the ray path will be determined by

$$\delta \int \frac{ds}{c + lu + mv + nw} = 0,$$

or, to the first order in $(u, v, w)/c$,

$$\delta \int \frac{ds}{c} - \delta \int \frac{1}{c^2}(u\,dx + v\,dy + w\,dz) = 0.$$

If $u\,dx + v\,dy + w\,dz$ is a perfect differential the second integral will depend solely on the values of u, v, w at the terminal points and will therefore be independent of the motion of the medium in between. The condition that $u\,dx + v\,dy + w\,dz$ should be a perfect differential is the condition that the motion of the medium should be what is known in hydrodynamics as differentially irrotational*. It means that there is to be no whirling motion. Provided this condition is satisfied the path of the ray passing between two points whose velocities are given is determined solely by the values of those velocities and does not depend on the motion of the medium in between.

As an illustration of Stokes's result we may consider the particular case of light propagated from a fixed point x_1, y_1, z_1 to a fixed point x_2, y_2, z_2, the space between being filled with a uniform stream of aether flowing with uniform velocity having components u, v, w. The path is determined by

---

* Cf. Lamb's *Hydrodynamics*, 1st ed., chap. III.

$$\delta\int_{x_1y_1z_1}^{x_2y_2z_2}\frac{ds}{c}-\delta\int_{x_1y_1z_1}^{x_2y_2z_2}\frac{1}{c^2}(u\,dx+v\,dy+w\,dz)=0,$$

or
$$\delta\int_{x_1y_1z_1}^{x_2y_2z_2}ds-\delta\left[\frac{1}{c}\{u(x_2-x_1)+v(y_2-y_1)+w(z_2-z_1)\}\right]=0,$$

or
$$\delta\int_{x_1y_1z_1}^{x_2y_2z_2}ds=0,$$

so that the path s is obviously a straight line from x_1, y_1, z_1 to x_2, y_2, z_2, just as if there were no motion.

The condition that the distribution of velocity communicated to the aether, by the motion of the earth through it, should be differentially irrotational is that the aether should behave like a perfect fluid for slow disturbances, such as the motion of material bodies through it would give rise to. This seems a natural condition of affairs, so that thus far Stokes was able to give a satisfactory account of aberration and still retain the view that the earth carried the aether in its immediate neighbourhood along with it, in the manner of a solid moving in an ordinary material fluid.

If the aether is an incompressible fluid it is not possible for it to be at rest relative to the surface of the earth and to have a velocity equal and opposite to that of the earth at distant points if its motion is continuous and irrotational*. Ways in which this difficulty could be overcome have been pointed out by Lorentz† and Planck‡. They seem, however, rather artificial. In order to explain Fizeau's experiment it is necessary, on Stokes's theory as well as Fresnel's, to suppose that a moving refracting medium imparts a velocity equal to $(1-m^{-2})$ of its own velocity to the aether within it. If the earth carries the aether with it in its immediate neighbourhood, as Stokes supposed, other moving bodies would be expected to have the same effect. An experiment to test this point has been made by Lodge§, who tried to find a change in the velocity of light in the region around the periphery of a massive iron flywheel when it was made to rotate rapidly. The results were entirely negative. A way of escape from most of the objections to Stokes's theory has recently been suggested by H. A. Wilson[*] (see p. 283).

---

* See Whittaker, *History of Theories of the Aether*, p. 412.

† *Arch. Neérl.* vol. XXI. p. 103 (1896).

‡ Cf. Lorentz, *Proc. Amsterdam Acad.* vol. I. p. 443 (1899).

§ *Phil. Trans.* A, vol. CLXXXIV. p. 727 (1893).

The Michelson and Morley Experiment.

So far those who held the view of an aether which is undisturbed by the motion of matter through it appeared to have much the better of the argument. The foregoing account represents the state of the question when Michelson and Morley carried out the famous experiment by which they hoped to obtain positive evidence of the relative motion of the earth and the aether. Before describing this experiment we shall first consider the problem of the reflection of light at a moving surface.

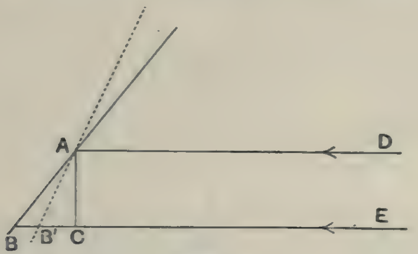


Fig. 38.

The plane mirror AB is moving towards the right with the uniform velocity v. A plane wave of light bounded by the rays DA, EB falls on it. Let us find the relation between the angle of incidence ι and the angle of reflection. When a wave meets the mirror at A on the ray DA, it has reached the point C on the ray EB. If the mirror were at rest this wave would afterwards meet it at B, but as the mirror is moving, the wave will meet it at B', where $BB'/B'C = v/c$, c being the velocity of light. The mirror thus behaves as though it were turned through an angle ϵ given by

$$\frac{\tan(\iota-\epsilon)}{\tan\iota} = \frac{B'C}{BC} = \frac{c}{c+v} = \frac{1}{\tan\iota}\,\frac{\tan\iota-\tan\epsilon}{1+\tan\iota\tan\epsilon},$$

---

[*] *Phil. Mag.* VI. vol. xix. p. 809 (1910).

whence
$$\frac{v}{c} = \tan\epsilon\left(\frac{1}{\tan\iota} + \tan\iota\right),$$

neglecting $\frac{\tan^2\epsilon}{\tan^2\iota}$. Hence to the first order in ϵ/ι

$$\epsilon = \tan\epsilon = \frac{1}{2}\frac{v}{c}\sin 2\iota.$$

The effect of the mirror on the reflected wave is therefore the same as if it were rotated through an angle $\epsilon = \frac{1}{2}\frac{v}{V}\sin 2\iota$. The reflected light will be turned, relatively to a similarly situated fixed mirror, through twice this angle.

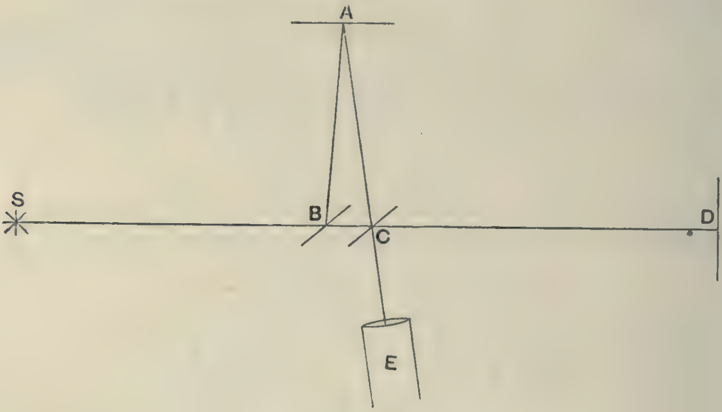


Fig. 39.

The arrangement of apparatus used in the Michelson and Morley experiment is shown in the accompanying diagram. Light from a source S passes through a half-silvered mirror B inclined at 45° to the direction of propagation. The reflected ray is reflected back by the plane mirror A so as to pass through the half-silvered mirror and thus reach the observing telescope E. The transmitted ray is reflected back by the plane mirror D and again by the half-silvered mirror, and so it also reaches the telescope E. The paths of the rays in space will depend on the magnitude and direction of the velocity of the apparatus. The figure has been drawn for the case in which the apparatus is carried along by the earth's motion in the direction of the ray SD. Let the velocity of the earth in its orbit be v, that of light in the undisturbed aether being c. Then the ray BA will not be perpendicular to SB, on account of the motion of the mirror, but will be inclined to this direction at an angle which is equal, to the first order, to $\theta = \tan^{-1} v/c$, by the preceding theorem, since $\iota = \pi/4$. Since A is moving parallel to itself, AC will be inclined to the normal at an equal angle on the other side of it. In an exactly similar way the ray CE produced by the reflection of CD at the moving mirror will be inclined at the same angle to CD as the ray AC. The two rays will therefore coincide in direction when they reach the observing telescope. If $AB = BD = l$ when the apparatus is at rest, the construction when it is in motion will be as shown, if C represents the position that B has moved into while the light moves from B to D and back. BC is evidently equal to $2vl/c$. The position of the interference fringes seen in the telescope will depend on the difference of time taken by the light to reach it along the two rays from B. This is equal to the difference of time along the two paths from B to C. If T_1 is the time along the path BAC this is given by

$$cT_1 = 2l\sqrt{\left(1 + \frac{v^2}{c^2}\right)}.$$

If T is the time required for the other ray to go from B to D, we have

$$cT = l + vT,$$

and if T' is the time required for it to get back from D to C

$$cT' = l - vT'.$$

So that

$$T = \frac{l}{c - v},$$

$$T' = \frac{l}{c + v},$$

and

$$\begin{aligned} T + T' - T_1 &= \frac{l}{c}\left\{\left(1 - \frac{v}{c}\right)^{-1} + \left(1 + \frac{v}{c}\right)^{-1} - 2\left(1 + \frac{v^2}{c^2}\right)^{\frac{1}{2}}\right\} \\ &= \frac{l}{c}\left\{1 + \frac{v}{c} + \frac{v^2}{c^2} + \dots + 1 - \frac{v}{c} + \frac{v^2}{c^2} + \dots \right. \\ &\qquad\qquad \left. - 2\left(1 + \frac{1}{2}\frac{v^2}{c^2} - \dots\right)\right\} \\ &= \frac{l}{c}\frac{v^2}{c^2}, \text{ neglecting higher powers of } \frac{v}{c}. \end{aligned}$$

The corresponding difference of path is

$$c(T+T'-T_1)=l\frac{v^2}{c^2}.$$

If the fringes were adjusted when the apparatus was stationary a displacement corresponding to this path difference should be observed if the whole apparatus were moving with uniform velocity v in the direction indicated. If the motion were with uniform velocity in a direction perpendicular to this the path difference is found to be equal but in the opposite direction; thus if the apparatus is adjusted so that SD coincides with the earth's motion relative to the sun, the effect of turning it through a right angle should be to cause the fringes to shift by an amount corresponding to a path difference equal to $2l\frac{v^2}{c^2}$.

Michelson and Morley set up their apparatus on a stone slab which floated in mercury. It could thus be rotated without causing strains to be set up in the apparatus. As the rotation was made to take place the fringes were observed continuously but no displacement of their position could be detected. The investigators were able to show conclusively that the relative motion of their apparatus and the aether could not amount to as much as one sixth part of the velocity of the earth in its orbit.

The result of the Michelson and Morley experiment was to place the problem of the relative motion of matter and aether in an apparently impossible position. On the one hand the view maintained by Stokes that the aether was carried along by the earth in its motion appeared to be full of inconsistencies, on the other hand the consequences of the stagnant aether hypothesis were flatly contradicted by experiment. An escape from the dilemma was pointed out by Fitzgerald who suggested that the null effect in the Michelson and Morley experiment was due to a change in the dimensions of the apparatus in the direction of the earth's motion, just sufficient to counterbalance the expected effect. The required change would evidently be produced if the matter of the apparatus contracted in the direction of the earth's motion in the ratio $1-v^2/2c^2$ to 1, the lengths in planes perpendicular to this direction being unaltered. This change would be too small to be capable of direct measurement in any case; but even if it were not, the operation could not be carried out; as any material scale which might be used would also contract, in the same ratio as the material to be measured.

This hypothesis seems a wild speculation at first sight, but it appears, on further inquiry, that it is rather what might be expected to occur if matter is made up of electrons. In that case the question at once suggests itself as to why a given portion of solid matter preserves its shape. If the matter were made up of superposable elements of positive and negative electrification, capable of infinitesimal subdivision, the only state of stable equilibrium would be one in which any excess of charge of one sign would be dissipated to infinity and the remaining equal and opposite charges would be superposed on each other. The matter would thus annihilate itself and disappear. In order that matter should be stable enough to preserve its identity it is necessary that the ultimate elements of electrification should be finite, and it is also necessary that the superposition of opposite elements should not result in annihilation. To ensure this it is necessary that the ultimate elements of opposite sign should not have exactly equal geometrical distributions. We conclude, in fact, that the positive and negative electrons are essentially different. The simplest assumption we can make as to the nature of the forces which keep them in equilibrium is to suppose that they are under the influence of equal and opposite forces of ultimately electrical character, but it may be that this supposition is inadequate and that the electrical forces are balanced by forces of non-electrical type. In any event, in static cases the equilibrium configuration of the electrons will be determined by the positions in the matter at which the resultant force is zero.

We know from the results of Chapter XI that the field due to an electric system in motion differs from that due to the same system at rest, in such a way as would result if all the lengths parallel to the direction of motion were changed in the ratio of $(1-v^2/c^2)^{\frac{1}{2}}$ to 1. Thus for the electrical forces to have the same value in the moving as in the fixed system it is necessary that all lengths in the former which are parallel to the direction of motion should be reduced in the ratio of $(1 - v^2/c^2)^{\frac{1}{2}}$ to 1. If the forces are all electrical it follows that the positions of the balancing points where the force vanishes will all be nearer together in this ratio and a contraction in length of this amount, in the direction of motion, will have to occur, in the moving matter, if equilibrium is to be maintained. Since

$$(1 - v^2/c^2)^{\frac{1}{2}} = 1 - v^2/2c^2$$

as far as the fourth order in v/c, we see that this contraction is of the right magnitude to account for the absence of an effect in the Michelson and Morley experiment. Another mode of explanation will be developed later.

An attempt has been made by Trouton and Rankine (*R. S. Proc.* A, vol. LXXX. p. 420, 1908) to detect and measure the Fitzgerald shrinkage by measuring the resistance of a metal strip when (*a*) parallel, and (*b*) perpendicular, to the earth's motion. The experiment showed that if any shrinkage occurs it is compensated to within 2 per cent. of the expected value by some counterbalancing effect of the motion on the resistance. Trouton and Rankine were able to show that on the electron theory of metallic conduction the changes in the mass, mean free path and velocity of the electrons would exactly compensate the effect of the expected shrinkage. It is interesting to note that according to their calculation the Lorentz change of electromagnetic mass with velocity gives the desired compensation, whereas the value found by Abraham does not.

Besides those which have already been discussed, there are a number of other cases, where effects due to motion through the aether might be expected to arise, which have engaged the attention of physicists. As is well known, the phenomenon of ordinary double refraction can be fully accounted for solely by the fact that the media which exhibit it transmit the disturbance which constitutes light with different velocities in different directions. We have seen that in order to explain Arago's experimental result it is necessary to suppose that the velocity of light relative to the observer, through a refracting medium which is isotropic when at rest, depends on the motion of the latter through the aether. The elative velocity of the wave will therefore be different in different lirections, and we might expect double refraction to occur when he medium is moving. Such an effect has been very carefully ought for but always with negative results.

Another effect which has been looked for is a change in the otation of the plane of polarization of plane polarized light proluced by its passage through a quartz plate, due to change in the elative direction of the apparatus and the earth's motion. The otation of the plane of polarization of light produced by its assage through quartz in a direction near that of the optic axis s believed to be due to the difference of velocity of circularly olarized rays when travelling in that direction. Thus the quartz s able to increase the velocity of one of the two oppositely lirected circularly polarized rays, into which the plane polarized ay may be resolved, relatively to the other. The emergent ray, roduced by their recombination, is thus polarized in a plane lifferent from that of the incident ray. Since the velocity of ither of the circularly polarized rays in the moving quartz lepends on a function of its refractive index for that ray multilied by the velocity of the quartz through the aether, and the efractive index is different for the two rays, an effect should arise hich is proportional to v/c. Experiment shows that there is no ffect which is comparable with that to be expected on this view. he reader will find a number of other experiments and observaions, which bear upon the subject under discussion, described n the last chapter of Whittaker's *History of the Aether and lectricity.*

A method of reconciling Stokes's theory of the optical proerties of moving bodies with the experimental facts, without ssuming the Fitzgerald contraction, has recently been indicated y H. A. Wilson[*]. Wilson points out that the problem of the notion of a body like the earth, through the aether, may be solved, n such a way that the motion of the incompressible aether is verywhere continuous and irrotational, provided that the tanential relative velocity vanishes at the surface of the body. All he conditions may be satisfied by an appropriate flow of the aether along the normal to the surface of the body. All the experiments to detect relative aether flow which have been made deal only with the velocity component parallel to the earth's surface and, owing to strains, etc., caused by the gravitational action, it would be almost impossible to execute delicate tests for motion in the vertical direction.

---

[*] *Phil. Mag.* VI. vol. xix. p. 809 (1910).

Relativity of Effects.

In reviewing the outcome of the experiments which have been undertaken with the object of discovering the relation between the motion of matter and aether, the most striking feature appears to be the small number of experiments which have led to a positive effect. The only cases where motion of matter appears to influence optical phenomena are

1. Astronomical aberration;
2. Relative motion of a refracting medium as in Fizeau's experiment; and
3. The Doppler effect.

It can hardly be a coincidence that all these cases entail the relative motion of matter. There is no experimental evidence of any optical effects arising from the motion of matter relative to aether or to space. One is therefore tempted to inquire whether it is really necessary to postulate an aether for the propagation of optical and electrical effects; it may appear that a more consistent set of relations would be obtained solely by the relative motion of matter. We shall have more to say about this question later.

The Propagation of Light in a Moving Refracting Medium.

The effect of the motion of a refracting medium on the propagation of light through it may be calculated on the electron theory, as follows:—

We shall consider the case of a plane polarized beam of light which is propagated along the z axis.

Referring to Chapter VIII, p. 148, we see that the polarization is related to the electromotive intensity E' by the equation

$$P = (m^2 - 1) E'.$$

Vhen the medium is in motion, with velocity V,

$$E' = E + \frac{1}{c} [V.H],$$

f the magnetic permeability of the medium is unity, in accordance vith equation (5), Chapter IX. In the present case, therefore,

$$P_x = (m^2 - 1) \left(E_x - \frac{w}{c} H_y \right) \quad \text{..............(2)},$$

f we take the axis of x parallel to the electric intensity and hat of y parallel to the magnetic intensity and the motion of the nedium parallel to the axis of z with velocity w. The current lensity is $\dot{D}_x + w \frac{\partial P_x}{\partial z}$, the latter term arising from the convection f the polarization by the moving medium. Thus the fundanental electromagnetic equations become

$$-\frac{\partial E_x}{\partial z} = \frac{1}{c} \dot{H}_y \quad \text{......................(3)},$$

nd
$$-\frac{\partial H_y}{\partial z} = \frac{1}{c} \left(\dot{D}_x + w \frac{\partial P_x}{\partial z} \right) = \frac{1}{c} \left(\dot{E}_x + \dot{P}_x + w \frac{\partial P_x}{\partial z} \right) \quad \text{...(4)}.$$

Eliminating P and H from (4) by means of (2) and (3) we have

$$\left(1 - (m - 1) \frac{w^2}{c^2} \right) \frac{\partial^2 E_x}{\partial z^2} = \frac{m^2}{c^2} \frac{\partial^2 E_x}{\partial t^2} + 2 \frac{w}{c^2} (m^2 - 1) \frac{\partial^2 E_x}{\partial t \partial z}.$$

If V is the velocity of light in the moving dielectric with espect to the fixed aether we shall have $E_x = A e^{ip(t - z/V)}$, where A and p are constants. Substituting this value and neglecting w^2/c^2 we have

$$c^2 = m^2 V^2 - 2w (m^2 - 1) V,$$

or
$$V = \frac{c}{m} + \left(1 - \frac{1}{m^2} \right) w \quad \text{...................(5)}.$$

This is Fresnel's formula (p. 273). This deduction shows that the convection of the polarized dielectric through the fixed aether produces the same change in the velocity of the light as is required by the experiments of Airy and Fizeau and by Fresnel's hypothesis. Those experiments therefore do not prove that the aether is carried along by the moving matter. The substance of this calculation was first given by Lorentz* and somewhat later by Larmor†.

Moving Axes and Correlated Systems.

We have seen that the behaviour of an electromagnetic system at the earth's surface may be represented by the system of differential equations:—

$$\operatorname{div} E = \rho \quad \text{(I)},$$

$$\operatorname{div} H = 0 \quad \text{(II)},$$

$$\operatorname{rot} H = (\dot{E} + \rho V)/c \quad \text{(III)},$$

$$\operatorname{rot} E = -\dot{H}/c \quad \text{(IV)},$$

where the force F on a unit charge is given by

$$F = E + [V \cdot H]/c \quad \text{(V)},$$

and V is the velocity of an element of charge relative to the earth's surface. If matter has a purely electrical constitution this system of equations will also describe the changes which material systems undergo.

A method, for which we are indebted to H. A. Lorentz, of investigating the effect on electrodynamic actions of the motion of the medium, is to transform the equations so that they refer to axes moving relatively to the material system instead of being fixed in it. As an illustration, and to fix our ideas, let us suppose that the above equations are true when the coordinates x, y, z and t are measured along axes fixed in the aether. Then $V = (u_1, v_1, w_1)$ will be the velocity of an element of charge relative to the fixed axes. We shall now examine how the differential equations are changed when they are made to refer to axes moving with uniform velocity u along the positive direction of the x axis, relative to the origin of the fixed axes Ox, Oy, Oz. Let the coordinates referred to the moving axes be denoted by x', y', z'; we shall also distinguish the time for events described with reference to the moving system by t'. Let the two systems of axes coincide at $t = 0$. Then

$$x' = x - ut, \quad y' = y, \quad z' = z, \quad t' = t.$$

---

* *Arch. Neérl.* vol. xxv. p. 525 (1892).

† *Phil. Trans.* A, vol. CLXXXV. p. 821 (1894).

x, y, z, t and x', y', z', t' being respective sets of independent variables we have

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x'}, \quad \frac{\partial}{\partial y} = \frac{\partial}{\partial y'}, \quad \frac{\partial}{\partial z} = \frac{\partial}{\partial z'},$$

but in virtue of the principle illustrated in dealing with the convection of the polarization charges on p. 285 we see that

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x}.$$

Thus $$\frac{\partial E_x}{\partial t} + u_1\rho = \frac{\partial E_x}{\partial t'} - u\frac{\partial E_x}{\partial x} + u_1\rho$$

$$= \frac{\partial E_x}{\partial t'} + (u_1 - u)\rho + u\left(\frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right),$$

from (I), and

$$\frac{\partial E_x}{\partial t'} + (u_1 - u)\rho = \frac{\partial}{\partial y'}(cH_z - uE_y) - \frac{\partial}{\partial z'}(cH_y + uE_z),$$

from (III).

Similarly $$\frac{\partial E_y}{\partial t'} + v_1\rho = \frac{\partial E_y}{\partial t} + u\frac{\partial E_y}{\partial x} + v_1\rho$$

$$= \frac{\partial}{\partial z'}(cH_x) - \frac{\partial}{\partial x'}(cH_z - uE_y),$$

and $$\frac{\partial E_z}{\partial t'} + w_1\rho = \frac{\partial}{\partial x'}(cH_y + uE_z) - \frac{\partial}{\partial y'}(cH_x).$$

Also $$-\frac{1}{c}\frac{\partial H_x}{\partial t'} = -\frac{1}{c}\left(\frac{\partial H_x}{\partial t} + u\frac{\partial H_x}{\partial x}\right)$$

$$= -\frac{1}{c}\frac{\partial H_x}{\partial t} + \frac{1}{c}\left(u\frac{\partial H_y}{\partial y} + u\frac{\partial H_z}{\partial z}\right).$$

Thus from (IV)

$$-\frac{\partial H_x}{\partial t'} = \frac{\partial}{\partial y'}(cE_z + uH_y) - \frac{\partial}{\partial z'}(cE_y - uH_z).$$

Similarly

$$-\frac{\partial H_y}{\partial t'} = \frac{\partial}{\partial z'}(cE_x) - \frac{\partial}{\partial x'}(cE_z + uH_y),$$

$$-\frac{\partial H_z}{\partial t'} = \frac{\partial}{\partial x'}(cE_y - uH_z) - \frac{\partial}{\partial y'}(cE_x).$$

Thus if we put

$$\left.\begin{aligned} H' &= (H_x', H_y', H_z') = \left(H_x,\ H_y + \frac{u}{c} E_z,\ H_z - \frac{u}{c} E_y\right) \\ E' &= (E_x', E_y', E_z') = \left(E_x,\ E_y - \frac{u}{c} H_z,\ E_z + \frac{u}{c} H_y\right) \\ V' &= (u', v', w') \qquad = (u_1 - u,\ v_1,\ w_1) \\ &\qquad\qquad \rho' = \rho \end{aligned}\right\} \ldots(6),$$

the equations may be written

$$\left.\begin{aligned} \frac{1}{c}\left(\frac{\partial E_x}{\partial t'} + u'\rho'\right) &= \frac{\partial H_z'}{\partial y'} - \frac{\partial H_y'}{\partial z'}, & -\frac{1}{c}\frac{\partial H_x}{\partial t'} &= \frac{\partial E_z'}{\partial y'} - \frac{\partial E_y'}{\partial z'} \\ \frac{1}{c}\left(\frac{\partial E_y}{\partial t'} + v'\rho'\right) &= \frac{\partial H_x'}{\partial z'} - \frac{\partial H_z'}{\partial x'}, & -\frac{1}{c}\frac{\partial H_y}{\partial t'} &= \frac{\partial E_x'}{\partial z'} - \frac{\partial E_z'}{\partial x'} \\ \frac{1}{c}\left(\frac{\partial E_z}{\partial t'} + w'\rho'\right) &= \frac{\partial H_y'}{\partial x'} - \frac{\partial H_x'}{\partial y'}, & -\frac{1}{c}\frac{\partial H_z}{\partial t'} &= \frac{\partial E_y'}{\partial x'} - \frac{\partial E_x'}{\partial y'} \end{aligned}\right\} \ldots(7).$$

The unaccented symbols in the left-hand members of (7) may be eliminated by means of the equations (6). We have, for example,

$$E_y' = E_y - \frac{u}{c} H_z = E_y\left(1 - \frac{u^2}{c^2}\right) - \frac{u}{c} H_z';$$

thus if

$$\beta^2 = (1 - u^2/c^2)^{-1},$$

$$E_y = \beta^2\left(E_y' + \frac{u}{c} H_z'\right),$$

and similarly

$$(H_x,\ H_y,\ H_z) = \beta^2\left(\beta^{-2} H_x',\ H_y' - \frac{u}{c} E_z',\ H_z' + \frac{u}{c} E_y'\right),$$

$$(E_x,\ E_y,\ E_z) = \beta^2\left(\beta^{-2} E_x',\ E_y' + \frac{u}{c} H_z',\ E_z' - \frac{u}{c} H_y'\right).$$

Substituting in (7) we find

$$\left.\begin{aligned} \frac{1}{c}\left(\frac{\partial E_x'}{\partial t'} + u'\rho'\right) &= \frac{\partial H_z'}{\partial y'} - \frac{\partial H_y'}{\partial z'} \\ \frac{1}{c}\left(\beta^2 \frac{\partial E_y'}{\partial t'} + v'\rho'\right) &= \frac{\partial H_x'}{\partial z'} - \left(\frac{\partial}{\partial x'} + \beta^2 \frac{u}{c^2} \frac{\partial}{\partial t'}\right) H_z' \\ \frac{1}{c}\left(\beta^2 \frac{\partial E_z'}{\partial t'} + w'\rho'\right) &= \left(\frac{\partial}{\partial x'} + \beta^2 \frac{u}{c^2} \frac{\partial}{\partial t'}\right) H_y' - \frac{\partial H_x'}{\partial y'} \\ -\frac{1}{c}\ \frac{\partial H_x'}{\partial t'} &= \frac{\partial E_z'}{\partial y'} - \frac{\partial E_y'}{\partial z'} \\ -\frac{1}{c}\beta^2 \frac{\partial H_y'}{\partial t'} &= \frac{\partial E_x'}{\partial z'} - \left(\frac{\partial}{\partial x'} + \beta^2 \frac{u}{c^2} \frac{\partial}{\partial t'}\right) E_z' \\ -\frac{1}{c}\beta^2 \frac{\partial H_z'}{\partial t'} &= \left(\frac{\partial}{\partial x'} + \beta^2 \frac{u}{c^2} \frac{\partial}{\partial t'}\right) E_y' - \frac{\partial E_x'}{\partial y'} \end{aligned}\right\} \ldots(8).$$

The differences between the values of H' and H and E' and E, respectively, required by equations (6), are in accordance with the experimental values of the magnetic force due to a changing electric field and the force on an electric charge moving in a magnetic field. These terms would therefore be expected to arise if the electromagnetic system under investigation were in motion relative to the measuring instruments employed to investigate it. Equations (8) show, however, that if rapid changes with time are occurring the state of the moving system will be different from that of the fixed system on account of the terms involving $\beta^2 \frac{u}{c^2} \frac{\partial}{\partial t'}$ and of terms of the second order in $\frac{u}{c}$ depending on β^2.

We have seen in Chap. XII that the effects due to a moving charge are not established instantaneously but are propagated through space with the constant velocity c which, we have reason to believe, is independent of the velocity of the matter at which the effects originate. It may be that on account of this finite velocity of propagation c the hypothesis ($t = t'$) that the time of an event when referred to the fixed axes is the same as that of the same event when referred to the moving axes, is leading to inconsistent results. Just as, according to Fitzgerald's hypothesis, the lengths of all material bodies change when the system is in motion, it may be that all the clocks change in a similarly definite manner. In such a case the local time t' might be a function of x, y, z as well as t. Without pursuing this discussion further for the present (we shall return to it later, p. 298), we shall provisionally admit the possibility that t' is different from t, and see to what conclusions we are led thereby.

In order to avoid altering the notation in the preceding equations we shall still retain $t' = t$ and introduce a new variable t'' for the local time. Following Lorentz, to whom the conception of local time is due, let us assume that

$$t'' = t' - \beta^2 \frac{u}{c^2} x'.$$

Then just as in changing from x to $x' = x - ut$ we have $\frac{\partial}{\partial x'} = \frac{\partial}{\partial x}$

and $\frac{\partial}{\partial t'} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x}$ we now have, in changing to the new variable t'',

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial t''} \text{ and } \frac{\partial}{\partial x''} = \frac{\partial}{\partial x'} + \beta^2 \frac{u}{c^2}\frac{\partial}{\partial t'}.$$

Thus the equations of the field become

$$\left.\begin{aligned}
\frac{1}{c}\left(\frac{\partial E_x'}{\partial t''} + u'\rho'\right) &= \frac{\partial H_z'}{\partial y'} - \frac{\partial H_y'}{\partial z'} \\
\frac{1}{c}\left(\beta^2\frac{\partial E_y'}{\partial t''} + v'\rho'\right) &= \frac{\partial H_x'}{\partial z'} - \frac{\partial H_z'}{\partial x''} \\
\frac{1}{c}\left(\beta^2\frac{\partial E_z'}{\partial t''} + w'\rho'\right) &= \frac{\partial H_y'}{\partial x''} - \frac{\partial H_x'}{\partial y'} \\
-\frac{1}{c}\quad\frac{\partial H_x'}{\partial t''} &= \frac{\partial E_z'}{\partial y'} - \frac{\partial E_y'}{\partial z'} \\
-\frac{1}{c}\beta^2\frac{\partial H_y'}{\partial t''} &= \frac{\partial E_x'}{\partial z'} - \frac{\partial E_z'}{\partial x''} \\
-\frac{1}{c}\beta^2\frac{\partial H_z'}{\partial t''} &= \frac{\partial E_y'}{\partial x''} - \frac{\partial E_x'}{\partial y'}
\end{aligned}\right\} \quad\text{.........(9).}$$

If we overlook the dashes these equations only differ from (III) and (IV) by the inclusion of the factor β^2, when it occurs on the left-hand side. Since $\beta^2 = (1 - u^2/c^2)^{-1}$ the two sets of equations are identical in form as far as the first order in u/c. This leads to an important principle which we have established to this order of accuracy.

Lorentz's Principle of Correlation.

If we can solve (III) and (IV) for any one of the variables E_x, E_y, E_z, H_x, H_y, H_z, let us say E_x, and then express the solution in the form

$$E_x = f(x, y, z, t),$$

this means that E_x, the x component of the electric intensity in the fixed system, is a certain function f of x, y, z and t, the space and time coordinates referred to the fixed axes. It follows from the identity up to the order $\frac{u}{c}$ of the equations (9) with (III) and (IV) that

$$E_x' = f(x'', y', z', t'').$$

is precisely the same function of

$$x'' = x - ut,\ y' = y,\ z' = z,\ t'' = t - \beta^2 \frac{u}{c} x''$$

that E_x is of x, y, z, t. Thus in comparing a moving and a fixed system, which are identical with one another when at relative rest, for any event in the fixed system characterised by a certain set of values of E and H, there is in the moving system a corresponding event characterised by values of E' and H', which are the same functions of x'', y', z' and t'' that E and H are of x, y, z and t. It follows that within a self-contained system, to the order of u/c, the electromagnetic effects are independent of the velocity of the system, since E' and H' are the values of the electric and magnetic forces which would be actually measured by instruments moving along with the system. This is only true, however, provided the time recorded by clocks in the moving system is the local time and not the "true" time recorded by clocks in the fixed system.

The principle of correlation is due to Lorentz. In the first instance he only succeeded in establishing it to the first order of u/c after the manner of the discussion above. A little later Larmor by including a contraction along the axes of motion, to accord with Fitzgerald's hypothesis, showed that the principle held to the order u^2/c^2. In 1903 Lorentz showed that a very similar transformation enabled the correlative principle to be established with exactness for all values of the velocity u less than that of light. We shall now investigate what is the necessary transformation of the electromagnetic equations in order to establish correspondence up to any order of u/c.

Since the differential equations are linear the transformation we are seeking will be a linear one and the preceding investigation at once suggests the form which is most likely to be successful. Let x, y, z, t be the coordinates referred to the first set of axes and x', y', z', t' those referred to the second, which are in motion relatively to the first. Assume that both x' and t' are linear functions of both x and t whilst y' and z' involve only y and z respectively. Let us put

$$x' = i(x + jt),\ y' = ky,\ z' = lz,\ t' = m(t + nx) \ldots\ldots (10),$$

where i, j, k, l, m, n are constants to be determined. Then

$$\left.\begin{aligned} i\frac{\partial}{\partial x'} &= \frac{\partial}{\partial x} - n\frac{\partial}{\partial t} = \frac{\partial}{\partial x} - mn\frac{\partial}{\partial t'} \\ k\frac{\partial}{\partial y'} &= \frac{\partial}{\partial y} \\ l\frac{\partial}{\partial z'} &= \frac{\partial}{\partial z} \\ m\frac{\partial}{\partial t'} &= \frac{\partial}{\partial t} - j\frac{\partial}{\partial x} = \frac{\partial}{\partial t} - ij\frac{\partial}{\partial x'} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(11).$$

Since the axes are moving in the x direction, $k = l$ by symmetry. We shall now transform the equations (III) and (IV) in which x, y, z and t are the independent variables, to new equations in terms of the independent variables x', y', z' and t'.

We have $$\frac{\partial E_x}{\partial t} + u_1\rho = m\frac{\partial E_x}{\partial t'} + ij\frac{\partial E_x}{\partial x'} + u_1\rho$$

$$= ck\left\{\frac{\partial H_z}{\partial y'} - \frac{\partial H_y}{\partial z'}\right\}.$$

Hence $$m\frac{\partial E_x}{\partial t'} + u_1\rho = ck\frac{\partial H_z}{\partial y'} - ck\frac{\partial H_y}{\partial z'} - ij\frac{\partial E_x}{\partial x'}.$$

Similarly $$\frac{\partial E_y}{\partial t} + v_1\rho = m\frac{\partial E_y}{\partial t'} + ij\frac{\partial E_y}{\partial x'} + v_1\rho$$

$$= ck\frac{\partial H_x}{\partial z'} - ci\frac{\partial H_z}{\partial x'} - cmn\frac{\partial H_z}{\partial t'}.$$

Thus

$$m\frac{\partial}{\partial t'}(E_y + cnH_z) + v_1\rho = ck\frac{\partial H_x}{\partial z'} - ci\frac{\partial}{\partial x'}\left(H_z + \frac{j}{c}E_y\right).$$

Similarly

$$m\frac{\partial}{\partial t'}(E_z - cnH_y) + w_1\rho = ci\frac{\partial}{\partial x'}\left(H_y - \frac{j}{c}E_z\right) - ck\frac{\partial H_x}{\partial y'}.$$

Also, from (I)

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho = i\frac{\partial E_x}{\partial x'} + mn\frac{\partial E_x}{\partial t'} + k\frac{\partial E_y}{\partial y'} + k\frac{\partial E_z}{\partial z'}.$$

Hence $$i\frac{\partial E_x}{\partial x'} = \rho - mn\frac{\partial E_x}{\partial t'} - k\frac{\partial E_y}{\partial y'} - k\frac{\partial E_z}{\partial z'}.$$

Hence the three equations may be written

$$\left.\begin{aligned}
m(1-nj)\frac{\partial E_x}{\partial t'}+(u_1+j)\rho &= ck\frac{\partial}{\partial y'}\left(H_z+\frac{j}{c}E_y\right)-ck\frac{\partial}{\partial z'}\left(H_y-\frac{j}{c}E_z\right)\\
m\frac{\partial}{\partial t'}(E_y+cnH_z)+v_1\rho &= ck\frac{\partial H_x}{\partial z'}-ci\frac{\partial}{\partial x'}\left(H_z+\frac{j}{c}E_y\right)\\
m\frac{\partial}{\partial t'}(E_z-cnH_y)+w_1\rho &= ci\frac{\partial}{\partial x'}\left(H_y-\frac{j}{c}E_z\right)-ck\frac{\partial H_x}{\partial y'}
\end{aligned}\right\}\quad\ldots\ldots(12).$$

Turning to the other three equations

$$-\frac{1}{c}\left(m\frac{\partial H_x}{\partial t'}+ij\frac{\partial H_x}{\partial x'}\right)=-\frac{1}{c}\frac{\partial H_x}{\partial t}=k\left(\frac{\partial E_z}{\partial y'}-\frac{\partial E_y}{\partial z'}\right),$$

and from (II)

$$\frac{\partial H_x}{\partial x}+\frac{\partial H_y}{\partial y}+\frac{\partial H_z}{\partial z}=0=i\frac{\partial H_x}{\partial x'}+mn\frac{\partial H_x}{\partial t'}+k\frac{\partial H_y}{\partial y'}+k\frac{\partial H_z}{\partial z'}.$$

Hence

$$\left.\begin{aligned}
-m(1-nj)\frac{\partial H_x}{\partial t'} &= ck\frac{\partial}{\partial y'}\left(E_z-\frac{j}{c}H_y\right)-ck\frac{\partial}{\partial z'}\left(E_y+\frac{j}{c}H_z\right)\\
-m\frac{\partial}{\partial t'}(H_y-cnE_z) &= ck\frac{\partial E_x}{\partial z'}-ci\frac{\partial}{\partial x'}\left(E_z-\frac{j}{c}H_y\right)\\
-m\frac{\partial}{\partial t'}(H_z+cnE_y) &= ci\frac{\partial}{\partial x'}\left(E_y+\frac{j}{c}H_z\right)-ck\frac{\partial E_x}{\partial y'}
\end{aligned}\right\}\quad\ldots\ldots(13).$$

The problem is to determine the values of the constants $i \ldots n$ and the functions $E_x' \ldots H_z'$ which will make equations (12) and (13) take the same forms as (III) and (IV). One obvious requirement is that terms such as $H_z+\frac{j}{c}E_y$ and H_z+cnE_y should be identical. Hence $n=j/c^2$. Making this substitution we have

$$\left.\begin{aligned}
m\left(1-\frac{j^2}{c^2}\right)\frac{\partial E_x}{\partial t'}+(u_1+j)\rho &= ck\frac{\partial}{\partial y'}\left(H_z+\frac{j}{c}E_y\right)-ck\frac{\partial}{\partial z'}\left(H_y-\frac{j}{c}E_z\right)\\
m\frac{\partial}{\partial t'}\left(E_y+\frac{j}{c}H_z\right)+v_1\rho &= ck\frac{\partial H_x}{\partial z'}-ci\frac{\partial}{\partial x'}\left(H_z+\frac{j}{c}E_y\right)\\
m\frac{\partial}{\partial t'}\left(E_z-\frac{j}{c}H_y\right)+w_1\rho &= ci\frac{\partial}{\partial x'}\left(H_y-\frac{j}{c}E_z\right)-ck\frac{\partial H_x}{\partial y'}
\end{aligned}\right\}\quad\ldots\ldots(14),$$

$$\left.\begin{aligned}
-m\left(1-\frac{j^2}{c^2}\right)\frac{\partial H_x}{\partial t'} &= ck\frac{\partial}{\partial y'}\left(E_z-\frac{j}{c}H_y\right)-ck\frac{\partial}{\partial z'}\left(E_y+\frac{j}{c}H_z\right)\\
-m\frac{\partial}{\partial t'}\left(H_y-\frac{j}{c}E_z\right) &= ck\frac{\partial E_x}{\partial z'}-ci\frac{\partial}{\partial x'}\left(E_z-\frac{j}{c}H_y\right)\\
-m\frac{\partial}{\partial t'}\left(H_z+\frac{j}{c}E_y\right) &= ci\frac{\partial}{\partial x'}\left(E_y+\frac{j}{c}H_z\right)-ck\frac{\partial E_x}{\partial y'}
\end{aligned}\right\}\quad\ldots\ldots(15).$$

Now put $j=-u,\ k=1,$

and $i=m=1\Big/\left\{m\left(1-\frac{j^2}{c^2}\right)\right\}=1\Big/\left(1-\frac{j^2}{c^2}\right)^{\frac{1}{2}}=1\Big/\left(1-\frac{u^2}{c^2}\right)^{\frac{1}{2}}=\beta.$

Then

$$\left.\begin{aligned}
\frac{1}{c}\left(\frac{\partial E_x}{\partial t'}+(u_1-u)\beta\rho\right) &= \frac{\partial}{\partial y'}\left\{\beta\left(H_z-\frac{u}{c}E_y\right)\right\}-\frac{\partial}{\partial z'}\left\{\beta\left(H_y+\frac{u}{c}E_z\right)\right\}\\
\frac{1}{c}\left(\frac{\partial}{\partial t'}\left\{\beta\left(E_y-\frac{u}{c}H_z\right)\right\}+v_1\rho\right) &= \frac{\partial H_x}{\partial z'}-\frac{\partial}{\partial x'}\left\{\beta\left(H_z-\frac{u}{c}E_y\right)\right\}\\
\frac{1}{c}\left(\frac{\partial}{\partial t'}\left\{\beta\left(E_z+\frac{u}{c}H_y\right)\right\}+w_1\rho\right) &= \frac{\partial}{\partial x'}\left\{\beta\left(H_y+\frac{u}{c}E_z\right)\right\}-\frac{\partial H_x}{\partial y'}
\end{aligned}\right\}\quad\ldots\ldots(16),$$

$$\left.\begin{aligned}
-\frac{1}{c}\frac{\partial H_x}{\partial t'} &= \frac{\partial}{\partial y'}\left\{\beta\left(E_z+\frac{u}{c}H_y\right)\right\}-\frac{\partial}{\partial z'}\left\{\beta\left(E_y-\frac{u}{c}H_z\right)\right\}\\
-\frac{1}{c}\frac{\partial}{\partial t'}\left\{\beta\left(H_y+\frac{u}{c}E_z\right)\right\} &= \frac{\partial E_x}{\partial z'}-\frac{\partial}{\partial x'}\left\{\beta\left(E_z+\frac{u}{c}H_y\right)\right\}\\
-\frac{1}{c}\frac{\partial}{\partial t'}\left\{\beta\left(H_z-\frac{u}{c}E_y\right)\right\} &= \frac{\partial}{\partial x'}\left\{\beta\left(E_y-\frac{u}{c}H_z\right)\right\}-\frac{\partial E_x}{\partial y'}
\end{aligned}\right\}\ldots(17).$$

Thus if we put

$$E_x'=E_x,\quad E_y'=\beta\left(E_y-\frac{u}{c}H_z\right),\quad E_z'=\beta\left(E_z+\frac{u}{c}H_y\right),$$

$$H_x'=H_x,\quad H_y'=\beta\left(H_y+\frac{u}{c}E_z\right),\quad H_z'=\beta\left(H_z-\frac{u}{c}E_y\right),$$

$$u'=u_1-u,\quad v'=v_1/\beta,\quad w'=w_1/\beta,\quad \rho'=\beta\rho,$$

the equations for the new dependent variables in terms of the independent variables $x'=\beta(x-ut)$, $y'=y$, $z'=z$ and $t'=\beta(t-ux/c^2)$ are accurately identical with the differential equations (III) and (IV) connecting the corresponding undashed variables. It follows that the correlation previously established to the first order is true to any order provided the electromagnetic quantities have the values given above and the time and space coordinates in the second system are the functions given above of those in the first system. The space in the second system is contracted in the direction of motion by a factor $(1 - u^2/c^2)^{\frac{1}{2}}$ as compared with that in the first system, whilst the time is contracted by a similar factor and, in addition, is a linear function of the x distance.

We shall postpone the further consideration of the physical consequences of this result, which was first given by Lorentz, until we have considered another method, which we owe to Einstein, of arriving at the same transformation.

CHAPTER XIV

THE PRINCIPLE OF RELATIVITY

As is well known, the Newtonian equations of motion retain their original form when the space and time coordinates are changed from x, y, z and t to $x' = x - vt$, $y' = y$, $z' = z$ and $t' = t$ corresponding to a uniform translatory motion of the axes with velocity v in the x direction. We have seen that this is no longer the case with the fundamental equations of electrodynamics. For the original form of the differential equations to be retained it is necessary that the x' coordinates should undergo a uniform contraction given by $x' = \beta(x - vt)$ and that t' should be similarly modified and also depend in a linear manner on x, being in fact given by $t' = \beta(t - vx/c^2)$. If therefore the units of space and time have the same values in two systems moving relatively to one another we should expect differences in similar phenomena occurring in them depending, at any rate, on the square of the ratio of their relative velocity to the velocity of light. All the experiments which have been made, however, lead to the conclusion that the actions taking place in any system depend only on the relative velocities of the parts of that system and are independent of the velocity relative to any other system which the system may have as a whole. We shall now attempt to see if, by changing our mode of defining time, we cannot make all phenomena independent of the state of motion of the system as a whole in which they occur. So far as matter consists of electrons and the phenomena in question are electromagnetic the problem has been solved in the last chapter; but the following treatment, due mainly to Einstein, is instructive and leads to results which are easier of application to many important problems.

In order to describe any physical phenomenon it is necessary to locate it in time and space. To locate it in space we must have three undeformable axes of reference and an equally rigid unit of length. We may then fix the time of the occurrence by having a series of similar clocks distributed at infinitesimal distances apart in space and arranged so that when brought together they register equal times. So far we have not indicated any method by which the clocks may be correlated when they are not in the same place. We shall now assume that this can be done in such a way that the velocity of light in a vacuum measured by means of the system of axes and the clocks is equal to a universal constant c, which is independent of the state of rest or of uniform motion of the system. Thus if A and B are two material points whose distance apart is r, a ray of light emitted from A at the instant t_A will reach B at the instant t_B, where

$$r = c(t_B - t_A);$$

t_A and t_B are the readings of the clocks at A and B respectively. We shall assume, what is not of course *a priori* obvious, that this is true whatever the state of motion of the axes of reference may be, provided it is not accelerated. The assumption here involved has been called by Einstein the Principle of the Constancy of the Velocity of Light. It is a particular case of the Principle of Relativity which may be stated in the words: The laws of nature are independent of the state of motion of the system of reference provided this is unaccelerated.

Now consider two equivalent systems of axes S and S'. By equivalent we mean here such as possess equal units of length and clocks running at equal rates, when the two systems are at rest relatively to each other. Suppose that S' is in uniform motion relatively to S. It follows from the fundamental assumption that the velocity of light must have the same value when measured with reference to both S and S'. Consider any event occurring at the point x, y, z at the time t when referred to the system S. Referred to the system S' this event will be described by the corresponding variables x', y', z' and t'. It is required to determine the relations between x, y, z and t and x', y', z' and t'.

On account of the homogeneity of time and space we should expect these relations to be linear. It follows that the coordinate planes of S' will be in uniform motion when referred to those of S. In general they will not be perpendicular to each other. Let us, however, choose the direction of the x and x' axes so that it coincides with the direction of relative motion of S and S'. Then by symmetry the coordinate planes of S' when referred to S will remain parallel to one another. We shall also choose the origin of time and the directions of the y and z axes so that the two systems S and S' coincide at the instant $t = 0$. Then the two following sets of equations have identical meanings:

$$\begin{aligned} x - vt &= 0 & \quad & x' = 0, \\ y &= 0 & \text{and} \quad & y' = 0, \\ z &= 0 & & z' = 0. \end{aligned}$$

Three of the desired transformations are therefore of the form

$$x' = a_1(x - vt), \qquad y' = b_1 y, \qquad z' = c_1 z.$$

Since the velocity of light has the same value c in reference to both systems and since they coincide at $t = t' = 0$ the following equations also have identical meanings:

$$x^2 + y^2 + z^2 = c^2 t^2,$$

and

$$x'^2 + y'^2 + z'^2 = c^2 t'^2.$$

Substituting in the last equation we have

$$\frac{a_1^2}{b_1^2}(x^2 - 2vxt + v^2t^2) + y^2 + c_1^2\frac{z^2}{b_1^2} = c^2\frac{t'^2}{b_1^2}.$$

On comparing with the first we see that $b_1^2 = c_1^2$, since x, y and z are independent and there is nothing to distinguish between y and z. This is obvious and we might have written $b_1 = c_1$ originally. Hence

$$\frac{a_1^2}{b_1^2}(x^2 - 2vxt + v^2t^2) - x^2 = \frac{c^2}{b_1^2}t'^2 - c^2t^2$$

must be an identity. From the homogeneity of this equation t' must be of the form $\alpha(t + \beta x)$, where α and β are constants. Hence

$$\frac{a_1^2}{b_1^2}(x^2 - 2vxt + v^2t^2) - x^2 + c^2t^2 - \frac{c^2}{b_1^2}(\alpha^2t^2 + 2\alpha^2\beta xt + \alpha^2\beta^2x^2) = 0,$$

for all values of t and x.

Hence

$$\frac{a_1^2}{b_1^2} - 1 - \frac{c^2}{b_1^2}\alpha^2\beta^2 = 0 \quad \text{.....................(1)},$$

$$-2v\frac{a_1^2}{b_1^2} - 2\alpha^2\beta\frac{c^2}{b_1^2} = 0 \quad \text{.....................(2)},$$

$$\frac{a_1^2}{b_1^2} v^2 + c^2 - \frac{c^2}{b_1^2} \alpha^2 = 0 \quad \text{....................} (3).$$

So that $\alpha^2 \beta c^2 = -va_1^2$ from (2).

Eliminating a_1 and α from (1) and (3) we find

$$1 + \beta v + \frac{\beta}{v} c^2 + c^2 \beta^2 = 0,$$

so that

$$\beta = -\frac{v}{c^2},$$

since the other root $\beta = -v^{-1}$ leads to inconsistent results, and

$$a_1^2 = \alpha^2 = \frac{b_1^2}{1 + \beta v} = \frac{b_1^2}{1 - v^2/c^2}.$$

Now b_1 cannot involve either x, y, z or t. It must therefore be a function of v only. We may denote it by $\phi(v)$. We have therefore determined all the coefficients except $\phi(v)$. Our results so far may be written

$$t' = \phi(v) \beta \left(t - \frac{v}{c^2} x\right), \qquad y' = \phi(v) y,$$

$$x' = \phi(v) \beta (x - vt), \qquad z' = \phi(v) z,$$

where β is now written for $1 \Big/ \sqrt{1 - \frac{v^2}{c^2}}$.

To determine $\phi(v)$ consider a third system of reference S'' similar to S and S' but moving relatively to S' with the velocity $-v$ along the x axis. For the time t'' referred to S'' we evidently have

$$\begin{aligned} t'' &= \phi(-v) \beta \left(t' + \frac{v}{c^2} x'\right) \\ &= \phi(-v) \beta \left[\phi(v) \beta \left\{t - \frac{v}{c^2} x + \frac{v}{c^2} (x - vt)\right\}\right] \\ &= \phi(-v) \phi(v) \beta^2 \left(1 - \frac{v^2}{c^2}\right) t \\ &= \phi(v) \phi(-v) t. \end{aligned}$$

Similarly

$$\begin{aligned} x'' &= \phi(v) \phi(-v) x, \\ y'' &= \phi(v) \phi(-v) y, \\ z'' &= \phi(v) \phi(-v) z. \end{aligned}$$

But since the systems S and S'' are always coincident, t'', etc. are identical with t, etc. So that

$$\phi(v) \phi(-v) = 1.$$

Moreover, since the relation between y and y' or z and z' cannot depend on the sign of v,

$$\phi(v) = \phi(-v),$$

and therefore $\phi(v) = 1$, $\phi(v) = -1$ being obviously untrue.

We therefore obtain finally

$$\left.\begin{aligned} t' &= \beta\left(t - \frac{v}{c^2}x\right) \\ x' &= \beta(x - vt) \\ y' &= y, \quad z' = z \end{aligned}\right\} \quad \text{......................(4).}$$

It will at once be observed that these values are the same as those which we obtained in the last chapter by the direct transformation of the electromagnetic equations.

If these equations are solved for t in terms of t', etc., we obtain

$$\left.\begin{aligned} t &= \beta\left(t' + \frac{v}{c^2}x'\right) \\ x &= \beta(x' + vt') \\ y &= y', \quad z = z' \end{aligned}\right\} \quad \text{........................(5),}$$

showing that the system S is moving with respect to the system S' with velocity $-v$ along the axis of x'.

It follows from these equations that when a body originally at rest is set in motion, its dimensions, measured relative to axes in the original position of rest, are contracted in the direction of motion and unchanged in planes perpendicular to this. Let x_1', y_1', z_1' and x_2', y_2', z_2' be the coordinates of any two points in the body referred to axes moving with it (system S'), x_1, y_1, z_1 and x_2, y_2, z_2 being the coordinates of the same point referred to the system S relative to which the body is moving parallel to the x axis with velocity v. Then at any time t which is constant with reference to the system S we have

$$\left.\begin{aligned} x_1 - x_2 &= (1 - v^2/c^2)^{\frac{1}{2}}(x_1' - x_2') \\ y_1 - y_2 &= y_1' - y_2' \\ z_1 - z_2 &= z_1' - z_2' \end{aligned}\right\} \quad \text{..............(6).}$$

Thus a length equal to l' parallel to the axis of x' is reduced in the ratio $\sqrt{1 - \frac{v^2}{c^2}} : 1$ when measured with reference to axes with regard to which the body is moving with velocity v at a constant time t referred to these axes.

If instead of t constant we take t' the time referred to the axes moving with the body as constant, then

$$x_1' - x_2' = (1 - v^2/c^2)^{\frac{1}{2}} (x_1 - x_2) \quad \text{.................} (7).$$

It follows that a body at rest with respect to S as measured from S' undergoes the same change of dimensions as a body at rest with respect to S' when measured from S. This result is required of course by the symmetry of the motion.

A similar set of relations holds with respect to the time. Suppose we have a clock moving with uniform velocity v along the axis of x when referred to the system S. Let us take the position of this clock as the origin of coordinates for the system S'. Then $x' = 0$ always and $x = vt$. Let t_1', t_2' be the times of two consecutive events as recorded on the moving clock and t_1, t_2 the times for the same events as registered in the fixed system S. Then

$$t_1' = \beta\left(t_1 - \frac{vx_1}{c^2}\right) = \beta t_1 \left(1 - \frac{v^2}{c^2}\right),$$

so that $$t_1 = \beta t_1' \text{ and } t_2 = \beta t_2'.$$

Hence $$t_2 - t_1 = \beta (t_2' - t_1') \quad \text{...................} (8).$$

We may take t_2', t_1' to represent two consecutive strokes of the clock. It is then clear from the equation above that the moving clock as observed from the fixed system will appear to have its periodic time increased in the ratio $1 : \sqrt{1 - \frac{v^2}{c^2}}$. The frequency will be decreased in the inverse ratio.

This case may be realised physically by considering the line spectrum emitted by a moving molecule or ion. Measured with reference to axes at rest with respect to the ion the frequency of the emitted light in the case of many lines is confined within very narrow limits. The period of the light may thus be taken to represent that of the reference clock and t_2', t_1', etc. may be taken as the times at which the emitted vibrations are consecutively in the same phase. The above result shows that in addition to the well-known Doppler effect the frequency ν of the light given out by an ion or molecule moving with velocity v relative to the axes of observation will be less than ν' that observed with reference to axes at rest relatively to the ion or molecule in accordance with the equation $\nu = \sqrt{1 - \frac{v^2}{c^2}}\,\nu'$.

Addition of Velocities.

Let any point move with a velocity, having components v_x', v_y', v_z', relative to the system of axes S', then

$(x' - x_0') = v_x'(t' - t_0')$, $(y' - y_0') = v_y'(t' - t_0')$, $(z' - z_0') = v_z'(t' - t_0')$.

Substituting for x', etc. in terms of x, etc. from equations (4) we find for the velocity components referred to the system S the values

$$\left.\begin{aligned} v_x &= \frac{x - x_0}{t - t_0} = \frac{v_x' + v}{1 + \frac{v_x' v}{c^2}} \\ v_y &= \frac{y - y_0}{t - t_0} = \frac{v_y'}{\beta\left(1 + \frac{v_x' v}{c^2}\right)} \\ v_z &= \frac{z - z_0}{t - t_0} = \frac{v_z'}{\beta\left(1 + \frac{v_x' v}{c^2}\right)} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(9).$$

It follows that the parallelogram of velocities is only true as a first approximation. If we put

$$v_1^2 = v_x^2 + v_y^2 + v_z^2,$$

and $$v_1'^2 = v_x'^2 + v_y'^2 + v_z'^2,$$

and let the angle between the x' axis measured with reference to the system S' and v_1' be denoted by α,

$$v_1 = \frac{\sqrt{(v^2 + v_1'^2 + 2vv_1'\cos\alpha) - \left(\frac{vv_1'\sin\alpha}{c^2}\right)^2}}{1 + \frac{vv_1'\cos\alpha}{c^2}}.$$

If v and v_1' are in the same direction

$$v_1 = \frac{v + v_1'}{1 + vv_1'/c^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10).$$

It follows from this equation that the resultant of two velocities each of which is less than that of light is also less than the velocity of light. For let $v = c - \lambda$ and $v_1' = c - \mu$, where λ and μ are positive and smaller than c, then the resultant velocity

$$v_1 = c\,\frac{2c - \lambda - \mu}{2c - \lambda - \mu + \lambda\mu/c},$$

and this is always less than c. Also the addition of any velocity to the velocity of light gives rise to a velocity which is still equal to the velocity of light.

Another interesting consequence of the foregoing results is that it is impossible for any signal to be transmitted from one point A to another B with a velocity W greater than c the velocity of light. Let the point A be taken at the origin of coordinates and the point B lie on the x axis at a distance l from A. Observers transmitting and receiving the signal are fixed at A and B respectively. Let the signal be transmitted by means of a material strip relatively to which it travels with velocity W in the direction $A \rightarrow B$. Now let the material strip carrying the signal be itself moving along the x axis with the velocity $-v$, where $v < c$, the velocity of light. The velocity of the signal relative to the transmitting and the observing system is evidently

$$\frac{W - v}{1 - Wv/c^2}.$$

The time required for the signal to be transmitted is thus

$$T = l\,\frac{1 - Wv/c^2}{W - v}.$$

Since v can have any value $< c$, T can always be made negative provided $W > c$. This would imply that the effect would be perceived before the cause had commenced to act. Although this may not necessarily involve any logical contradiction it is opposed to the whole character of our experience. The truth of the theorem therefore follows.

Some of the preceding results differ so considerably from those which follow from the generally accepted notions of space and time that many readers will probably regard them as serious objections to the views here developed. If, however, the principle of relativity is accepted they appear to follow with logical certainty. The *a priori* argument in favour of the principle of relativity will be considered later.

Application to Optics.

Let the vectors which describe a wave of light when referred to the system S be proportional to

$$\begin{matrix}\sin\\ \cos\end{matrix}\,\omega\left(t-\frac{lx+my+nz}{c}\right),$$

l, m, n being the direction cosines referred to S of the directions of propagation. Referred to the system S' let the vectors be proportional to

$$\begin{matrix}\sin\\ \cos\end{matrix}\,\omega'\left(t'-\frac{l'x'+m'y'+n'z'}{c}\right).$$

It follows from the principle of relativity that if t', x', y', z' are replaced by their values in terms of t, x, y, z drawn from equations (4) the two functions of t, x, y, z will be identically equal. Thus

$$\begin{aligned}&\begin{matrix}\sin\\ \cos\end{matrix}\,\omega'\left(t'-\frac{l'x'+m'y'+n'z'}{c}\right)\\ &\qquad=\begin{matrix}\sin\\ \cos\end{matrix}\,\omega'\left[\beta\left(1+\frac{l'v}{c}\right)t-\frac{\beta\left(l'+\frac{v}{c}\right)x+m'y+n'z}{c}\right]\\ &\qquad=\begin{matrix}\sin\\ \cos\end{matrix}\,\omega\left(t-\frac{lx+my+nz}{c}\right)\text{ identically.}\end{aligned}$$

Hence

$$\omega=\beta\left(1+l'\frac{v}{c}\right)\omega',\qquad m\omega=m'\omega',$$

$$l\omega=\beta\left(l'+\frac{v}{c}\right)\omega',\qquad n\omega=n'\omega'.$$

So that

$$\left.\begin{aligned}l'&=\frac{l-\frac{v}{c}}{1-\frac{v}{c}\,l}, & m'&=\frac{m}{\beta\left(1-\frac{v}{c}\,l\right)}\\ n'&=\frac{n}{\beta\left(1-\frac{v}{c}\,l\right)}, & \omega'&=\beta\left(1-l\frac{v}{c}\right)\omega\end{aligned}\right\}\quad\ldots\ldots(11).$$

The last formula is the complete expression for the Doppler effect. If an observer is moving with velocity v relative to a source of light (system S), so that *referred to the system S*

(i.e. the system of coordinates at rest with reference to the source) the velocity v makes an angle ϕ with the line joining the source and the observer, the frequency ν' of the light perceived by the observer is related to the frequency of the light as measured by reference to an observing system at rest with respect to the source by the relation

$$\frac{\nu'}{\nu} = \frac{\omega'}{\omega} = \frac{1 - \frac{v}{c}\cos\phi}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(12).$$

If on the other hand the angle is given with reference to the system S' moving with the observer, and is denoted by ϕ', we have

$$\frac{\nu'}{\nu} = \beta\left(1 - \frac{v}{c}\,l\right) = \beta\left(1 - \frac{v}{c}\,\frac{l' + v/c}{1 + l'\,v/c}\right)$$

$$= \frac{1}{\beta}\,\frac{1}{1 + l'\,\frac{v}{c}} = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v}{c}\cos\phi'} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(13).$$

If we regard the source of light as being in motion and the observer as fixed and if the frequency of the light when measured with reference to axes moving with the source is ν_0, if also the source is moving relative to the observer with velocity v making an angle ϕ with the line joining the source and the observer *when referred to axes at rest with reference to the observer*, then the frequency ν of the light perceived by the observer is given by

$$\frac{\nu}{\nu_0} = \frac{\omega}{\omega'}, = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{v}{c}\cos\phi}.$$

If our reasoning is correct this formula should also be given by substituting $-v$ for v and ϕ' for ϕ in the first equation for ν'/ν. For the change from moving observer to moving source is equivalent to a change of sign in the relative velocity; and the axes to which ϕ *was* referred have now become the moving axes. We thus get

$$\frac{\nu'}{\nu} = \beta\left(1 - l\frac{v}{c}\right) = \beta\left(1 + \frac{v}{c}\,l'\right) = \beta\left(1 + \frac{v}{c}\,\frac{l' - v/c}{1 - l'\,v/c}\right)$$

$$= \frac{1}{\beta}\,\frac{1}{1 - l'\,v/c},$$

which agrees with the last formula if $\nu' = \nu$, $\nu = \nu_0$ and $l' = \cos\phi$.

Aberration.

If the relative velocity v is parallel to the x axis and if ϕ is the angle between the direction of the ray and that of the relative motion when referred to the S system of axes, and ϕ' the same quantity when referred to the S' system, then

$$\cos\phi' = l' = \frac{\cos\phi - \frac{v}{c}}{1 - \frac{v}{c}\cos\phi} \quad \ldots\ldots\ldots\ldots\ldots\ldots(14).$$

This formula embraces the whole of the theory of astronomical aberration. Its relation to the formulae previously obtained may be left as an exercise to the student.

Fizeau's Experiment.

Suppose the light is travelling in a moving medium. Let this be at rest with respect to S', the axes of x and x' being chosen so as to coincide with the direction of relative motion. Then the light vectors will be proportional to $\frac{\sin}{\cos}\,\omega'\left(t' - \frac{x'}{V'}\right)$ or to $\frac{\sin}{\cos}\,\omega\left(t - \frac{x}{V}\right)$ according to the axes of reference chosen. Since the system S is moving with velocity $-v$ when referred to S' we obtain

$$\omega = \beta\omega'\left(1 + \frac{v}{V'}\right),$$

$$\frac{\omega}{V} = \beta\,\frac{\omega'}{V'}\left(1 + \frac{V'v}{c^2}\right).$$

Hence
$$V = \frac{V' + v}{1 + \frac{V'v}{c^2}} = V'\left(1 + \frac{v}{V'}\right)\left(1 - \frac{V'v}{c^2}\right) \text{ approx.}$$

$$= V'\left\{1 + \left(\frac{1}{V'} - \frac{V'}{c^2}\right)v\right\} \text{ approx.}$$

$$= V' + \left(1 - \frac{1}{m^2}\right)v \text{ approx.} \ldots\ldots\ldots\ldots(15).$$

This gives Fizeau's result to the order of accuracy with which it is capable of being verified.

Electromagnetic Equations.

The transformation of the electromagnetic equations of Maxwell to variables referred to the system S' has already been made in the last chapter. The equations assume the form

$$\left.\begin{aligned}\frac{1}{c}\left(\frac{\partial E_x'}{\partial t'}+u'\rho'\right)&=\frac{\partial H_z'}{\partial y'}-\frac{\partial H_y'}{\partial z'}\\ \frac{1}{c}\left(\frac{\partial E_y'}{\partial t'}+v'\rho'\right)&=\frac{\partial H_x'}{\partial z'}-\frac{\partial H_z'}{\partial x'}\\ \frac{1}{c}\left(\frac{\partial E_z'}{\partial t'}+w'\rho'\right)&=\frac{\partial H_y'}{\partial x'}-\frac{\partial H_x'}{\partial y'}\end{aligned}\right\}\text{..............}(a),$$

$$\left.\begin{aligned}-\frac{1}{c}\frac{\partial H_x'}{\partial t'}&=\frac{\partial E_z'}{\partial y'}-\frac{\partial E_y'}{\partial z'}\\ -\frac{1}{c}\frac{\partial H_y'}{\partial t'}&=\frac{\partial E_x'}{\partial z'}-\frac{\partial E_z'}{\partial x'}\\ -\frac{1}{c}\frac{\partial H_z'}{\partial t'}&=\frac{\partial E_y'}{\partial x'}-\frac{\partial E_x'}{\partial y'}\end{aligned}\right\}\text{..............}(b),$$

where

$$E_x'=E_x,\qquad E_y'=\beta\left(E_y-\frac{u}{c}H_z\right),\qquad E_z'=\beta\left(E_z+\frac{u}{c}H_y\right)\text{...}(c),$$

$$H_x'=H_x,\qquad H_y'=\beta\left(H_y+\frac{u}{c}E_z\right),\qquad H_z'=\beta\left(H_z-\frac{u}{c}E_y\right)\text{...}(d),$$

$$u'=\frac{u_1-v}{1-\dfrac{u_1v}{c^2}},\qquad v'=\frac{v_1}{\beta\left(1-\dfrac{u_1v}{c^2}\right)},\qquad w'=\frac{w_1}{\beta\left(1-\dfrac{u_1v}{c^2}\right)}\text{...}(e),$$

$$\rho'=\beta\left(1-\frac{u_1v}{c^2}\right)\rho\text{ ....}(f).$$

Previously we wrote $u'=u_1-v$, $\rho'=\beta\rho$, etc., but the above values of the velocities relative to S' are the values required by the theory and are consistent with the equations in the last chapter, provided we put $\rho'=\beta\left(1-\frac{u_1v}{c^2}\right)\rho$. This value of ρ' is thus the value required by the principle of relativity.

Since according to the principle of relativity the physical laws are independent of the motion of the system as a whole it follows that $E'(E_x', E_y', E_z')$ is the value of the electric intensity in the system as measured by an observer who with his apparatus is moving with it. If the system is moving relatively to the observer the forces will have a different value from the above in accordance with equation (c). It follows that the so-called electromotive forces acting on a moving charge in a magnetic field are nothing but electric forces when measured by an observer moving with the charge. From this standpoint the distinction between electric and magnetic forces becomes indefinite. By a suitable change of moving axes either may be made to vanish, involving a corresponding change in the other. For instance with a uniformly moving point charge there are important magnetic forces if the motion is relative to the observer, but if the observer moves with the charge the forces are all electric.

It follows from equation (f) that if a body is at rest relative to S' its total charge e' measured in reference to the system S' is the same as its total charge e measured in reference to S *at a definite instant t referred to S.* For the total charge e' referred to S' is

$$e' = \iiint_{\tau'} \rho'\, dx'\, dy'\, dz'.$$

Now at any particular instant t referred to S it follows from equations (4) that

$$dx'\, dy'\, dz' = \beta\, dx\, dy\, dz.$$

Also since the body is at rest referred to S', $u' = 0$, and therefore $u_1 = v$. Hence

$$\rho' = \beta\left(1 - \frac{u_1 v}{c^2}\right)\rho = \beta\left(1 - \frac{v^2}{c^2}\right)\rho = \frac{\rho}{\beta}.$$

So that $$e' = \iiint_{\tau'} \rho'\, dx'\, dy'\, dz' = \iiint_{\tau} \rho\, dx\, dy\, dz = e.$$

It follows that if any material system is set in motion the magnitudes of the charges, as determined from axes at rest with reference to the initial state of motion of the system, are unchanged.

The equations (a)—(f) suffice for the solution of all problems in the electromagnetics and optics of moving systems in which accelerations do not play an important part. As an illustration we shall use them to determine the amplitude A' relative to the system S' of a wave of light which is specified relatively to S by the equations

$$E = E_0 \sin \phi ; \quad H = H_0 \sin \phi ; \quad \phi = \omega \left(t - \frac{lx + my + nz}{c} \right).$$

Let E_x, E_y, E_z denote the components of E and E_1, E_2, E_3 the components of E_0, with a similar notation for the components of H and H_0.

Referring to the equations on p. 307 we see that the values of the vectors referred to S' are

$$E_x' = E_1 \sin \phi', \qquad H_x' = H_1 \sin \phi',$$

$$E_y' = \beta \left(E_2 - \frac{v}{c} H_3 \right) \sin \phi', \quad H_y' = \beta \left(H_2 + \frac{v}{c} E_3 \right) \sin \phi',$$

$$E_z' = \beta \left(E_3 + \frac{v}{c} H_2 \right) \sin \phi', \quad H_z' = \beta \left(H_3 - \frac{v}{c} E_2 \right) \sin \phi',$$

$$\phi' = \omega' \left(t' - \frac{l'x' + m'y' + n'z'}{c} \right).$$

The values of ω', l', m', n' have already been determined in terms of ω, l, m, n and v, and the consequences which follow from them have been discussed.

Let us determine the amplitude A' referred to S' for the case in which the electric vibration in the wave referred to S is parallel to the z axis. Then in free space (or in any isotropic medium) the direction of the ray will be in the plane of xy, and, if θ is the angle between this direction and that of the x axis,

$$E_1 = 0, \qquad E_2 = 0, \qquad E_3 = A,$$

$$H_1 = - A \sin \theta, \qquad H_2 = - A \cos \theta, \qquad H_3 = 0.$$

Thus

$$E_x' = 0, \qquad E_y' = 0, \qquad E_z' = \beta \left(1 - \frac{v}{c} \cos \theta \right) A \sin \phi',$$

$$H_x' = - A \sin \theta \sin \phi', \quad H_y' = \beta \left(- \cos \theta + \frac{v}{c} \right) A \sin \phi', \quad H_z' = 0.$$

Therefore A', the amplitude referred to the system S', is given by

$$A' = A \frac{1 - \frac{v}{c} \cos \theta}{\sqrt{1 - \frac{v^2}{c^2}}},$$

since $$\sin^2\phi + \frac{\left(\cos\phi - \frac{v}{c}\right)^2}{1-\frac{v^2}{c^2}} = \left[\frac{\left(1-\frac{v}{c}\cos\phi\right)}{\sqrt{1-\frac{v^2}{c^2}}}\right]^2.$$

As the direction of the z axis is arbitrary this formula must be true generally.

Mechanics of an Electron or Material Point.

Let an electric charge e, of infinitesimal dimensions, move under the influence of an electromagnetic field. We shall assume, in accordance with the principle of relativity, that its equations of motion, referred to the system of axes S' with reference to which it is instantaneously at rest, are

$$\mu\frac{d^2x_0'}{dt_0'^2} = eE_x', \qquad \mu\frac{d^2y_0'}{dt_0'^2} = eE_y', \qquad \mu\frac{d^2z_0'}{dt_0'^2} = eE_z',$$

where μ *is a constant* which we shall call the mass of the electron. The suffix 0 is introduced to indicate that the moving point is instantaneously at rest with reference to the S' axes.

Let us now deduce the equations of motion of the point charge when referred to the system S relative to which S' is moving with velocity v. We have

$$dt_0' = \beta\left(dt_0 - \frac{v}{c^2}dx_0\right) = \beta\left(1-\frac{v\dot{x}_0}{c^2}\right)dt_0,$$

$$dx_0' = \beta\,(dx_0 - v\,dt_0) \quad = \beta\,(\dot{x}_0 - v)\,dt_0,$$

$$dy_0' = dy_0, \qquad dz_0' = dz_0.$$

Hence $$\frac{dx_0'}{dt_0'} = \frac{\dot{x}_0 - v}{1-\frac{v\dot{x}_0}{c^2}},$$

$$\frac{d^2x_0'}{dt_0'^2} = \frac{d}{dt_0'}\left(\frac{\dot{x}_0 - v}{1-\frac{v\dot{x}_0}{c^2}}\right) = \frac{1}{\beta\left(1-\frac{v\dot{x}_0}{c^2}\right)}\frac{d}{dt_0}\left(\frac{\dot{x}_0 - v}{1-\frac{v\dot{x}_0}{c^2}}\right)$$

$$= \frac{1}{\beta}\frac{\left(1-\frac{v\dot{x}_0}{c^2}\right)\ddot{x}_0 + \frac{v}{c^2}(\dot{x}_0 - v)\,\ddot{x}_0}{\left(1-\frac{v\dot{x}_0}{c^2}\right)^3}.$$

Similarly
$$\frac{dy_0'}{dt_0'} = \frac{1}{\beta\left(1-\dfrac{v\dot{x}_0}{c^2}\right)}\frac{dy_0}{dt_0},$$

$$\frac{d^2y_0'}{dt_0'^2} = \frac{1}{\beta\left(1-\dfrac{v\dot{x}_0}{c^2}\right)}\frac{1}{\beta}\frac{d}{dt_0}\left(\frac{\dot{y}_0}{1-\dfrac{v\dot{x}_0}{c^2}}\right)$$

$$= \frac{\left(1-\dfrac{v\dot{x}_0}{c^2}\right)\ddot{y}_0 + \dot{y}_0\dfrac{v}{c^2}\ddot{x}_0}{\beta^2\left(1-\dfrac{v\dot{x}_0}{c^2}\right)^3},$$

with a similar equation] for $\dfrac{d^2z_0'}{dt_0'^2}$.

If S' is instantaneously at rest with reference to the moving point $\dot{x}_0 = v$, $\dot{y}_0 = 0$, $\dot{z}_0 = 0$, so that

$$\begin{aligned}
\frac{d^2x_0'}{dt_0'^2} &= \sqrt{1-\frac{v^2}{c^2}}\,\frac{\ddot{x}_0}{(1-v^2/c^2)^2} = \beta^3\ddot{x}_0,\\
\frac{d^2y_0'}{dt_0'^2} &= \frac{\ddot{y}_0}{\beta^2(1-v^2/c^2)^2} = \beta^2\ddot{y}_0,\\
\frac{d^2z_0'}{dt_0'^2} &= \frac{\ddot{z}_0}{\beta^2(1-v^2/c^2)^2} = \beta^2\ddot{z}_0.
\end{aligned}$$

Referring to equation (c), p. 307, we therefore get

$$\begin{aligned}
\mu\beta^3\ddot{x}_0 &= eE_x,\\
\mu\beta\ddot{y}_0 &= e\left(E_y - \frac{v}{c}H_z\right),\\
\mu\beta\ddot{z}_0 &= e\left(E_z + \frac{v}{c}H_y\right).
\end{aligned}$$

These are the equations which hold for the instant when $\dot{x}_0 = v$, $\dot{y}_0 = 0$, $\dot{z}_0 = 0$. We may on the left-hand side replace v by $q = \sqrt{\dot{x}_0^2 + \dot{y}_0^2 + \dot{z}_0^2}$ and on the right-hand by $\dot{x}_0$. Leaving out the suffix 0 and adding the other terms $\dfrac{\dot{y}}{c}H_z$, $-\dfrac{\dot{z}}{c}H_y$, the above equations may be written as a particular case of the symmetrical equations

$$\left.\begin{aligned}
\frac{d}{dt}\left\{\frac{\mu\dot{x}}{\sqrt{1-q^2/c^2}}\right\} &= K_x\\
\frac{d}{dt}\left\{\frac{\mu\dot{y}}{\sqrt{1-q^2/c^2}}\right\} &= K_y\\
\frac{d}{dt}\left\{\frac{\mu\dot{z}}{\sqrt{1-q^2/c^2}}\right\} &= K_z
\end{aligned}\right\}\quad\ldots\ldots\ldots\ldots\ldots\ldots(16),$$

where
$$\left.\begin{aligned} K_x &= e\left[E_x + \frac{\dot{y}}{c}H_z - \frac{\dot{z}}{c}H_y\right] \\ K_y &= e\left[E_y + \frac{\dot{z}}{c}H_x - \frac{\dot{x}}{c}H_z\right] \\ K_z &= e\left[E_z + \frac{\dot{x}}{c}H_y - \frac{\dot{y}}{c}H_x\right] \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots(17).$$

μ is a constant coefficient in all three equations and since

$$\frac{d}{dt}\left\{\frac{\mu\dot{x}}{\sqrt{1-q^2/c^2}}\right\} = \frac{\mu}{(1-q^2/c^2)^{\frac{3}{2}}}\left\{\ddot{x}\left(1 - \frac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{c^2}\right) + \dot{x}\,\frac{\dot{x}\ddot{x}+\dot{y}\ddot{y}+\dot{z}\ddot{z}}{c^2}\right\},$$

with similar expressions for the terms in $\dot{y}$ and $\dot{z}$, it will be seen at once that equations (16) reduce to the equations on p. 310 for the particular case when $\dot{y} = \dot{z} = 0$. Since the equations (16) and (17) retain their form when transformed to any new set of axes at rest referred to the first but differently directed in space, if they are true for any one set they will be true for any other. But we have seen that they are true when $\dot{y} = \dot{z} = 0$; they are therefore true in general.

The vectors K_x, K_y, K_z we shall call the components of the force acting in the electron. When q^2/c^2 is negligible the equations of motion are identical with those of Newton; otherwise they are not.

We shall extend the scope of equations (16) so as to embrace the case in which the forces are of gravitational origin. As their applicability in this case is a sheer assumption they can only be regarded as a definition of force.

Energy and Momentum.

If we multiply each component of the universal equations (III) and (IV), Chap. XIII, p. 286, in turn by $E_x, E_y, \ldots H_z$, add them together and integrate over a space at whose boundaries the electric and magnetic forces vanish, we obtain

$$\int \rho\,(E_x u_1 + E_y v_1 + E_z w_1)\,d\tau + \frac{dL}{dt} = 0,$$

where $$L = \int \tfrac{1}{2}\left[(E_x^2 + E_y^2 + E_z^2) + \tfrac{1}{2}(H_x^2 + H_y^2 + H_z^2)\right] d\tau$$

is the electromagnetic energy of the space considered. If the electric density ρ is due to electrons of charge e, the integral is equivalent to $\Sigma\, e\,(E_x\dot{x} + E_y\dot{y} + E_z\dot{z})$, where E_x, E_y, E_z are the components of the external electric force (i.e. the part not arising from the electron itself) acting on the electron. Thus, if the conservation of energy is to hold, the rate of working of the electric field on the electron is $e(E_x\dot{x} + E_y\dot{y} + E_z\dot{z})$. It will be seen from equations (17) that this is equal to $K_x\dot{x} + K_y\dot{y} + K_z\dot{z}$. Thus the force as above defined has the same relation to the rate of working as in the Newtonian Mechanics.

The equations (16) may be written in the form

$$\frac{\mu}{\sqrt{1-q^2/c^2}}\,d\dot{x} + \tfrac{1}{2}\,\frac{\mu\dot{x}}{(1-q^2/c^2)^{\frac{3}{2}}}\,\frac{d(q^2)}{c^2} = K_x dt,$$

with similar expressions in y and z. Multiplying each of these in turn by $\dot{x}$, $\dot{y}$ and $\dot{z}$, and adding, we get

$$\begin{aligned}(\dot{x}K_x + \dot{y}K_y + \dot{z}K_z)dt &= \tfrac{1}{2}\,\frac{\mu d(q^2)}{(1-q^2/c^2)^{\frac{1}{2}}}\left(1 + \frac{q^2/c^2}{(1-q^2/c^2)}\right) \\ &= d\left(\frac{\mu c^2}{\sqrt{1-q^2/c^2}}\right).\end{aligned}$$

The kinetic energy being equal to the work done by the external forces will therefore be

$$\int(\dot{x}K_x + \dot{y}K_y + \dot{z}K_z)dt = \frac{\mu c^2}{\sqrt{1-q^2/c^2}} + \text{const.}$$

The constant will of course have to be determined by the initial conditions. If the kinetic energy was zero when $q = 0$ we should have const. $= -\mu c^2$. The value of the kinetic energy for the system whose velocity is q is therefore

$$\mu c^2\left(\frac{1}{\sqrt{1-q^2/c^2}} - 1\right) \quad \text{..................(18).}$$

For small values of q this is equal to $\frac{1}{2}\mu q^2$ in agreement with the value of ordinary mechanics.

If we multiply the second and third of equations (a) and (b), p. 307, in turn by H_z, $-H_y$, E_z and $-E_y$ and add, we obtain, as in Chapter X, omitting the dashes,

$$\begin{aligned}&\frac{d}{dt}\frac{1}{c}(H_zE_y - H_yE_z) + \rho\left(\frac{v_1}{c}H_z - \frac{w_1}{c}H_y\right) \\ &\quad = H_z\frac{\partial H_x}{\partial z} - H_z\frac{\partial H_z}{\partial x} - H_y\frac{\partial H_y}{\partial x} + H_y\frac{\partial H_x}{\partial y} \\ &\qquad + E_z\frac{\partial E_x}{\partial z} - E_z\frac{\partial E_z}{\partial x} - E_y\frac{\partial E_y}{\partial x} + E_y\frac{\partial E_x}{\partial y}\end{aligned}$$

$$= \tfrac{1}{2}\frac{\partial}{\partial x}(H_x^2 - H_y^2 - H_z^2) + \frac{\partial}{\partial y}(H_x H_y) + \frac{\partial}{\partial z}(H_x H_z)$$

$$- H_x\left(\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z}\right) + \tfrac{1}{2}\frac{\partial}{\partial x}(E_x^2 - E_y^2 - E_z^2)$$

$$+ \frac{\partial}{\partial y}(E_x E_y) + \frac{\partial}{\partial z}(E_x E_z) - E_x\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right).$$

Integrating throughout a closed space over the boundaries of which the forces vanish, and remembering that div $H = 0$ and div $E = \rho$, the whole of the right-hand side vanishes except

$$\int E_x \rho \, d\tau,$$

and we obtain

$$\frac{d}{dt}\int \frac{1}{c}(H_z E_y - H_y E_z)\, d\tau + \int \rho \left(E_x + \frac{v_1}{c} H_z - \frac{w_1}{c} H_y\right) d\tau = 0,$$

hence
$$\frac{d}{dt}\int \frac{1}{c}(H_z E_y - H_y E_z)\, d\tau + \Sigma K_x = 0,$$

so that

$$\frac{d}{dt}\left[\int \frac{1}{c}(H_z E_y - H_y E_z)\, d\tau + \Sigma \frac{\mu \dot{x}}{\sqrt{1 - q^2/c^2}}\right] = 0 \ldots\ldots(19),$$

with similar equations in y and z.

Since $\frac{1}{c}(H_z E_y - H_y E_z)$ is the x component of electromagnetic momentum per unit volume of the system, the equations above express the law of conservation of momentum if

$$\xi = \frac{\mu \dot{x}}{\sqrt{1 - q^2/c^2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(20)$$

is the momentum attributed to a point charge whose mass is μ. We have as in mechanics $\frac{\partial \xi}{\partial t} = K_x$.

These values of the kinetic energy and momentum enable the equations of motion of an electron to be written in the Hamiltonian form. The student who finds any difficulty with this may be referred to Einstein (*Jahrbuch der Radioaktivität und Elektronik*, vol. IV. p. 435 (1907)).

Experimental Test.

For small values of q the laws of motion deduced from the principle of relativity are identical with the Newtonian Laws, but this is no longer the case when q^2 is comparable with c^2. By observing the effect of externally applied forces on electrons moving with very high velocities we might expect to make a test of the principle of relativity. There are three functions of the velocity q relative to the observing apparatus which might under favourable conditions be capable of affording observations for this purpose. These are (1) the potential difference required to produce the velocity q, (2) the deflection of the path of the moving electron by a stationary electric field, and (3) the corresponding deflection produced by a magnetic field. If e is the charge on an electron, the potential difference V required to increase its velocity relative to the observing system from zero to q is given by the equation (see p. 313)

$$Ve = \mu c^2 \left[\frac{1}{\sqrt{1 - q^2/c^2}} - 1 \right] \text{ ....(21).}$$

For the electric and magnetic deflections, consider the case in which the direction of q is instantaneously along the x axis. If a magnetic force M along the y axis and an electric force Z along the z axis act on the electron, its equation of motion will be

$$\frac{d^2z}{dt^2} = \frac{e}{\mu} \sqrt{1 - q^2/c^2} \left(Z + \frac{q}{c^2} M \right).$$

The path is therefore curved in the xz plane and the instantaneous radius of curvature R is given by $q^2/R = \frac{d^2z}{dt^2}$. The electric and magnetic deflections are therefore measured respectively by

$$\frac{1/R}{Z} = \frac{e}{\mu} \frac{\sqrt{1 - q^2/c^2}}{q^2},$$

$$\frac{1/R}{M} = \frac{e}{\mu} \frac{\sqrt{1 - q^2/c^2}}{c^2 q}.$$

These deflections vary with q in exactly the same way as those calculated in Chapter XI, for Lorentz's contractile electron. They have been tested, in the manner already described, by the experiments of Kaufmann and Bucherer. As the results of these experiments support Lorentz's calculations they are equally in favour of the relativity theory.

So far it has not been possible to test the relation between V and q experimentally. With the radium rays used by Kaufmann V is not under the control of the experimenter; and with cathode rays it is not possible to get values of q high enough to make the difference between formula (21) and the usual formula $Ve = \frac{1}{2}\mu q^2$ capable of being measured.

The Inertia of Energy.

Consider a physical system surrounded by an imaginary non-material enclosure which is impervious to radiation. The object of this is to prevent the energy of the system from escaping in the radiant form. Let external electric forces whose components are X_e, Y_e, Z_e act on the system. These enable any desired amount of energy to be introduced into the system from outside. The energy gained by the system when referred to the system of reference S, in accordance with our previous results, will be given by the expression

$$\int dE = \int dt \int \rho \left(X_e u_1 + Y_e v_1 + Z_e w_1\right) d\tau,$$

where ρ is the density of the electricity at any point in the enclosed space. Let us transform this equation so that the right-hand side is referred to variables proper to the system S'. We have

$$t = \beta\left(t' + \frac{v}{c^2}x'\right), \quad x = \beta\left(x' + vt'\right), \quad y = y', \quad z = z',$$

and
$$t' = \beta\left(t - \frac{v}{c^2}x\right), \quad x' = \beta\left(x - vt\right).$$

Now,
$$u_1 = \frac{u' + v}{1 + \frac{u'v}{c^2}}, \quad v_1 = \frac{v'}{\beta\left(1 + \frac{u'v}{c^2}\right)}, \quad w_1 = \frac{w'}{\beta\left(1 + \frac{u'v}{c^2}\right)},$$

and
$$\rho = \frac{\rho'}{\beta\left(1 - \frac{vu_1}{c^2}\right)} = \beta\left(1 + \frac{u'v}{c^2}\right)\rho'.$$

Also
$$X = X', \qquad Y = \beta\left(Y' + \frac{v}{c}N'\right), \qquad Z = \beta\left(Z' - \frac{v}{c}M'\right).$$

So that $$\int dt \int \rho\,(X_e u_1 + Y_e v_1 + Z_e w_1)\,d\tau$$

$$= \int dt \int dx \int dy \int dz\, F(x, y, z, t)$$

$$= \int dt \int dx \iint dy' dz'\, F'(x', y', z', t')$$

$$= \frac{1}{\beta} \int dt \iiint d\tau'\, F'(x', y', z', t')^*$$

$$= \beta \int dt' \int \rho' (X_e' u' + Y_e' v' + Z_e' w')\, d\tau'$$

$$+ \beta v \int dt' \int \rho' \left(X_e' + \frac{v'}{c} N_e' - \frac{w'}{c} M_e' \right) d\tau'.$$

Since the principle of relativity must also apply to the system S' this may be written

$$\int dE = \beta \int dE' + \beta v \int [\Sigma K_x']\, dt' \quad \text{..............(22).}$$

Consider the case in which the motion of the system as a whole is such that it is at rest relative to S', and suppose further that the velocities of its parts relative to S' are so small that v^2/c^2 may be considered negligible. The centre of mass of the system is thus at rest relative to S' which, under the further condition postulated, can only be the case provided $\Sigma K_x' = 0$ for all values of t'. In spite of this $\int [\Sigma K_x']\, dt'$ will not necessarily vanish; for this integral is not taken between given values of t' but of t, so that in general the limits will involve x' as well as t'.

If, however, the external forces do not act except during the interval considered, the parts which would otherwise depend on x' also vanish. This statement will be made clearer in the sequel where the general case is considered. If the forces vanish entirely outside the time limits then we have

$$\int dE = \beta \int dE',$$

so that $$dE = \beta dE'.$$

We therefore conclude that the energy of a uniformly moving system which is not under the influence of external forces is a function of two variables, namely its energy E_0 relative to a system of axes moving with it and its velocity of translation v relative to the standard system of axes. Thus

---

* See Williamson's *Integral Calculus*, p. 320.

$$\frac{\partial E}{\partial E_0} = \frac{1}{\sqrt{1 - v^2/c^2}},$$

so that

$$E = \frac{E_0}{\sqrt{1 - v^2/c^2}} + \psi(v).$$

$\psi(v)$ is evidently the energy of the system when $E_0 = 0$. This has already been determined for the case of a material point (see p. 313), and was found to be

$$\mu c^2 \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right).$$

The value for the whole uniformly moving system will therefore be obtained by addition of all the masses and is equal to

$$c^2 \Sigma \mu \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right),$$

where Σ denotes summation over the whole system. The complete expression for E is therefore

$$E = \left(\Sigma \mu + \frac{E_0}{c^2} \right) \left(\frac{c^2}{\sqrt{1 - v^2/c^2}} \right) - c^2 \Sigma \mu \quad \ldots\ldots\ldots\ldots (23).$$

Comparing this with the formula for the case in which $E_0 = 0$, namely

$$E = \frac{c^2}{\sqrt{1 - v^2/c^2}} \Sigma \mu - c^2 \Sigma \mu,$$

we see that, so far as the part of the energy which depends on the velocity v is concerned, the effect of the energy E_0 is to increase the apparent mass of the system from $\Sigma \mu$ to $\Sigma \mu + E_0/c^2$. Thus an addition δE to the energy of any system will give rise to an increase $\delta E/c^2$ in its mass. It is only a short step from this result to the hypothesis that all mass is simply a manifestation of confined energy.

The question whether or not mass is simply a manifestation of confined energy is obviously a matter of the very utmost importance, and it is very desirable that it should be submitted to the test of experiment. The energy liberated in chemical actions is so small compared with the "dead" masses involved that it is hopeless to detect any change of mass thus arising, always supposing that weight as well as mass is proportional to $\mu + E/c^2$. In the case of radioactive change the matter is more hopeful. The decrease of mass of a system due to the loss of 1000 gm. calories is $4{\cdot}6 \times 10^{-11}$ gm. Now 1 gm. atom of radium in radioactive equilibrium evolves about $3{\cdot}024 \times 10^4$ gm. cals. per hour. Thus its diminution of mass per hour would be

$$\frac{3{\cdot}024 \times 10^4 \times 4{\cdot}2 \times 10^7}{9 \times 10^{20}} = 1{\cdot}4 \times 10^{-6} \text{ mgm.}$$

This would amount in a year to ·012 mgm. or in 100 years to 1·2 mgm. It would be worth the while of anyone who could afford the necessary capital to make observations of this character.

Momentum.

We have seen that ξ, the x component of momentum of a point charge, satisfies the relation $\frac{d\xi}{dt} = K_x$. Let us apply this result to the system, previously considered, which is surrounded by a closed boundary impervious to radiation and is subjected to the action of external electric forces X_e, Y_e, Z_e, etc. Then the total momentum gained by the system in the time during which the forces act will be

$$\begin{aligned}\int d\xi = \int K_x dt &= \int dt \int \rho \left(X_e + \frac{v_1}{c} N_e - \frac{w_1}{c} M_e \right) d\tau \\ &= \beta \int dt' \int \rho' \left(X_e' + \frac{v'}{c} N_e' - \frac{w'}{c} M_e' \right) d\tau' \\ &\qquad + \frac{\beta v}{c^2} \int dt' \int \rho' \left(X_e' u' + Y_e' v' + Z_e' w' \right) d\tau' \\ &= \beta \frac{v}{c^2} \int dE' + \beta \int [\Sigma K_x'] \, dt'.\end{aligned}$$

Let the system move as before, so that its centre of gravity remains at rest referred to S', then $\Sigma K_x' = 0$, and if in addition the forces are zero outside the time limits considered, the second integral vanishes and therefore

$$\int d\xi = \beta \frac{v}{c^2} \int dE'$$

and

$$d\xi = \beta \frac{v}{c^2} dE' \quad \text{......................(24).}$$

The momentum is therefore also a function of the energy referred to axes moving with the moving system and of the relative velocity q $(=v)$ of the latter referred to the standard system. We have in fact

$$\frac{\partial \xi}{\partial E_0} = \frac{v/c^2}{\sqrt{1 - v^2/c^2}},$$

so that
$$\xi = \frac{v}{\sqrt{1 - v^2/c^2}} \frac{E_0}{c^2} + \psi'(v).$$

$\psi'(v)$ is obviously the momentum when $E_0 = 0$. This has already been determined for a point charge, and it is clear that for the whole uniformly moving system

$$\psi'(v) = \frac{v \Sigma \mu}{\sqrt{1 - v^2/c^2}},$$

where $\Sigma\mu$ is the sum of the masses contained in the system. The complete expression for the momentum is thus

$$\xi = \left[\Sigma\mu + \frac{E_0}{c^2}\right] \frac{v}{\sqrt{1 - v^2/c^2}} \quad \text{...............(25).}$$

Thus the effect of the internal energy E_0 on the momentum of the system is equivalent to an increase of the mass by the amount E_0/c^2, in agreement with the change we have found it to produce in the energy itself.

Forces Continuously Operative.

Let us now consider the case referred to on p. 317 when the forces continue to act outside the limits of the time t, still confining ourselves to the case where the centre of gravity of the moving enclosed system remains at rest relative to S', so that for any instant t', $\Sigma K_x'$ is still equal to zero. Under these circumstances the integral $\int(\Sigma K_x')\,dt'$ would vanish if the limits of integration were given values of t', but as they are given values of t it does not. Let the time limits with respect to S be t_1 and t_2; then the limits for t' are determined by the equation (see p. 300)

$$t = \beta\left(t' + \frac{v}{c^2} x'\right).$$

They are therefore $\frac{t_1}{\beta} - \frac{v}{c^2} x'$ and $\frac{t_2}{\beta} - \frac{v}{c^2} x'$.

The limits for t' thus depend on the x' coordinate as well as on t. We may split the integral to be evaluated into three parts thus:

$$\int [\Sigma K_x']\, dt' = \left[\int_{\frac{t_1}{\beta}-\frac{v}{c^2}x'}^{\frac{t_1}{\beta}} + \int_{\frac{t_1}{\beta}}^{\frac{t_2}{\beta}} + \int_{\frac{t_2}{\beta}}^{\frac{t_2}{\beta}-\frac{v}{c^2}x'}\right] [\Sigma K_x']\, dt'.$$

The middle integral vanishes, since its limits do not depend on x'. We are not able to evaluate the other two integrals in the general case in which K_x' varies in an arbitrary manner with the time. However, the most important case practically is that in which the change of K_x' during a time comparable with vx'/c^2 is negligible, and in this case the integrals assume a very simple form. Under these conditions we have

$$\int_{\frac{t_1}{\beta}-\frac{v}{c^2}x'}^{\frac{t_1}{\beta}} (\Sigma K_x')\, dt' = \frac{v}{c^2} \Sigma K_x' x'.$$

The sum of the two integrals may evidently be written

$$\int [\Sigma K_x']\, dt' = + \left[\frac{v}{c^2} \Sigma K_x' x'\right]_{t_2}^{t_1}.$$

The calculation of the energy and momentum is now easily carried out in the manner previously employed, and we obtain

$$E = \left(\Sigma\mu + \frac{E_0}{c^2}\right) \frac{c^2}{\sqrt{1 - v^2/c^2}} - \frac{v^2/c^2}{\sqrt{1 - v^2/c^2}} \sum_{t_2} K_x' x' - c^2 \Sigma\mu \quad \text{...(26)},$$

$$\xi = \frac{v}{\sqrt{1 - v^2/c^2}} \left[\Sigma\mu + \frac{E_0 - \sum_{t_2} (K_x' x')}{c^2}\right] \quad \text{......................(27)}.$$

In these formulae K_x' is to be interpreted as the x' component of the external force and x' the x' coordinate of its point of application, both referred to the system S'. In the particular case where K_x' arises from a uniform external hydrostatic pressure acting on the boundary of the system, equal to p_0 when referred to the system of axes S', then if V_0 is the volume of the enclosed space referred to the same axes,

$$\sum_{t_2} (K_x' x') = - p_0 V_0,$$

and
$$E = \left(\Sigma\mu + \frac{E_0}{c^2}\right) \frac{c^2}{\sqrt{1 - v^2/c^2}} + \frac{v^2/c^2}{\sqrt{1 - v^2/c^2}} p_0 V_0 - c^2 \Sigma\mu \quad \text{...(28)},$$

$$\xi = \frac{v}{\sqrt{1 - v^2/c^2}} \left(\Sigma\mu + \frac{E_0 + p_0 V_0}{c^2}\right) \quad \text{........................(29)}.$$

Examples.

If the moving system consists of electromagnetic radiation enclosed in a massless closed boundary, the energy and momenta referred to axes relative to which the system is moving with velocity v are

$$E = \frac{E_0}{\sqrt{1 - v^2/c^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (30),$$

$$\xi = \frac{v}{\sqrt{1 - v^2/c^2}} \frac{E_0}{c^2} = \frac{v}{c^2} E \quad \ldots\ldots\ldots\ldots\ldots\ldots (31),$$

where E_0 is the energy of the system referred to axes moving with it.

These are the values on the assumption that the boundary walls are rigid. If, however, the boundary is perfectly flexible the radiation pressure will have to be balanced by an external pressure p_0 given by the relation

$$p_0 V_0 = \tfrac{1}{3} E_0/c^2.$$

In this case
$$E = \frac{E_0 (1 + \frac{1}{3} v^2/c^2)}{\sqrt{1 - v^2/c^2}},$$

$$\xi = \frac{v}{\sqrt{1 - v^2/c^2}} \frac{\frac{4}{3} E_0}{c^2}.$$

The principle of relativity leads to the conclusion that almost all physical quantities are functions of the velocity v of the system relative to the axes of reference. It would lead us beyond the scope of this book to go into the matter in detail, but the following table of corresponding values is instructive:

Physical Quantity	Value referred to axes moving with system	Value referred to system S
Pressure of hydrostatic type	p_0	$p = p_0$ ......(32),
Confined energy	E_0	$E = \beta E_0$ ......(33),
Temperature	T_0	$T = \frac{1}{\beta} T_0$ ......(34),
Entropy	η_0	$\eta = \eta_0$ ......(35).

The reader who is interested in this subject may be referred for further information to Planck ("Zur Dynamik bewegter Systeme," *Sitzungsber. d. kgl. Preuss. Akad. der Wissenschaften*, 1907) and Einstein (*Jahrbuch der Radioakt. u. Elektronik*, vol. IV. p. 451, 1907).

The Principle of Relativity and the Aether.

Before leaving this part of our subject it is desirable to review the bearing of the principle of relativity on the question of the existence of the luminiferous aether. We have seen that if we start from the hypothesis that electromagnetic actions have their seat in a medium, the aether, which is at absolute rest, the known facts can be explained on the hypothesis of Fitzgerald that bodies contract when in motion relative to the aether. This contraction can be shown to be a plausible consequence of the motion through the aether. If this contraction is the only change due to the motion, the effects in moving systems would not be exactly correlated with those in fixed systems, although the differences, so far as the writer is able to judge, would not have been detected in any experiments which have been carried out up to the present. Any effects which might arise which were not accounted for by such a scheme might be explained by making the velocity of light a function of the motion of the system through the aether.

On the other hand, the principle of relativity, which is in accordance with all the known facts, describes them in a simpler and more symmetrical manner. It is clear that if the principle of relativity and its consequences are valid, electromagnetic experiments can never yield any information as to the state of rest or uniform motion of an aether. This follows since, in the last analysis, all the effects are then made to depend on the relative motion of matter. It is, in fact, quite unnecessary ever to bring the word aether into the discussion.

From this standpoint it is desirable, perhaps, to state the matter somewhat more explicitly. To specify any physical event it is necessary to locate it in time and space, that is to say, to determine the four coordinates t, x, y and z of the time and place at which it occurs. The question arises as to whether we can be sure that two events which appear to occur at the same place at successive times t and t' really occur at the same place. Can we be sure that the place which appears to be the same has not changed its position in the interval? It is clear that such a discussion is futile until we have fixed on a set of axes Ox, Oy, Oz in terms of which we can specify the position of the points considered.

Let us suppose that we have fixed upon such a set of axes and that they are so chosen that the physical system in which the events occur is at rest as a whole when referred to these axes. Having marked off our axes in terms of measuring rods, which we may suppose to be part of the system, we should naturally turn our attention to the measurement of time. This could be done by means of a series of clocks which we could compare with one another at some fixed point, let us say at the origin. It would, however, be necessary to have some means of comparing them when they were moved away so as to be at a considerable distance from one another. This could be done by sending light signals. The simplest assumption we could make would be that the light was propagated in spherical waves. By considering the case of a wave propagated from the origin we see that x, y, z and t would satisfy the relation $x^2+y^2+z^2-c^2t^2=0$, where c is the velocity of light in space. The coordinates as thus determined would be consistent with each other and would be the simplest ones in terms of which the events in that system could be described. If we had no opportunity to investigate other systems we should probably conclude that our system was at rest relative to the medium in which the light was propagated.

Now suppose that we have another system, let us say a distant solar system, which is moving with the velocity v relative to the first. An investigator located in the second system would be able to discover a framework of axes and a set of coordinates x_1, y_1, z_1 and t_1 in terms of which light would in his system be propagated in spherical waves. He would find these coordinates the simplest in terms of which he could describe the events occurring in his own system, and he would have the same reason for concluding that the second system was at rest relative to the aether that the first observer had had for concluding that his was. It is evident that one of the two conclusions must be fallacious, and there is nothing to favour one rather than the other.

The coordinates x, y, z, t and x_1, y_1, z_1, t_1 which refer to a given event are of course different for the two systems. Each set is preferable to the other for describing events in its own system; but if we are to describe events in a universe in which both systems exist there is nothing to choose between them. It is clear that there are an infinite number of possible systems of reference corresponding to all the possible values of v, and each one is as good as another. In fact, if we wish to represent the whole universe in the simplest and most elegant manner, we cannot thus arbitrarily separate time and space, but we must rather consider the whole as a four-dimensional manifold of x, y, z, and t.

Four-dimensional analysis has now been devised with the object of effecting such a representation. It is found that the electromagnetic equations then assume a more symmetrical form than that in which we have considered them. The reader who is interested in these questions may be referred to the following authorities:—

Minkowski, *Raum und Zeit.* Leipzig, 1909.
Sommerfeld, *Ann. der Physik*, vol. XXXII. p. 749, and vol. XXXIII. p. 649. 1910.
M. Laue, *Das Relativitäts Princip.*

To sum up we may say that the Principle of Relativity furnishes no evidence either for or against the existence of an aether. It denies the possibility of determining the motion of such a fluid if it exists. In so far as Relativity is a Universal Principle it finds the aether a superfluous hypothesis.

CHAPTER XV

RADIATION AND TEMPERATURE

It is a very familiar fact that when material bodies are heated they emit electromagnetic radiations, in the form of thermal, luminous and actinic rays, in appreciable quantities. Such an effect is a natural consequence of the electron and kinetic theories of matter. On the kinetic theory, temperature is a measure of the violence of the motion of the ultimate particles; and we have seen that, on the electron theory, electromagnetic radiation is a consequence of their acceleration. The calculation of this emission from the standpoint of the electron theory alone is a very complex problem which takes us deeply into the structure of matter and which has probably not yet been satisfactorily resolved. Fortunately we can find out a great deal about these phenomena by the application of general principles like the conservation of energy and the second law of thermodynamics without considering special assumptions about the ultimate constitution of matter. It is to be borne in mind that the emission under consideration occurs at all temperatures although it is more marked the higher the temperature.

The problem which we set before us is that of finding the nature of the radiation which is found in an enclosure containing material bodies and maintained at a constant temperature. Sufficient time is supposed to have elapsed for any secular changes in the enclosed matter to have come to an end, so that the nature and condition of any element of the matter does not vary. Special radiations of chemical or radioactive origin are therefore eliminated in so far as they involve progressive material changes. Even in the steady state the interchange of radiation will be accompanied by an interchange of electrons arising from thermionic and photoelectric emission (see Chap. XVIII). The effect of this on the calculations may be eliminated by special devices, for instance by surrounding the radiating surfaces by an envelope of ideal matter which is perfectly transparent to radiation but perfectly opaque to electrons, and as it makes no difference in the final results we shall, for the sake of brevity, leave it out of account.

In considering a train of plane electromagnetic waves, the intensity is defined as the amount of energy which they transport in unit time across unit area of a surface perpendicular to their direction of propagation. If the normal to a surface is inclined at an angle θ to the direction of propagation, the amount of energy which it would receive per unit area, in unit time, is equal to the intensity multiplied by $\cos\theta$. In the case of the radiation in an enclosure the specification of the intensity is not so simple; for here the wave trains are travelling indiscriminately in all directions. Let $d\omega$ be an element of solid angle and dS an element of area described normally about the axis of $d\omega$. If $i(\omega)\,dS\,d\omega$ is the energy of the radiation, incident on dS in unit time, whose direction of propagation is comprised within the element of solid angle $d\omega$, then $i(\omega)$ is the intensity of this radiation. The notation $i(\omega)$ is employed to indicate the possibility that this quantity may depend upon the direction of the axis of $d\omega$. We shall see that one of the characteristic properties of the radiation present in an enclosure maintained at a constant temperature is that $i(\omega)$ is independent of this direction. It is clear that if the normal to dS is inclined at an angle θ to the axis of $d\omega$, the energy incident on dS in unit time which is propagated in directions lying within $d\omega$ is

$$i(\omega)\cos\theta\,dS\,d\omega.$$

The radiation under consideration involves other elements in its composition beside the solid angle $d\omega$. We know that by means of a prism or other analysing device it can be split into elements having different frequencies (ν). The result of this analysis is independent of the instrument used provided the latter does not absorb or transform the energy in any way. The process is, in fact, a physical resolution into the equivalent Fourier's series. It is therefore legitimate to express i as an infinite sum of terms, or as an equivalent definite integral, extending over all the possible frequencies between 0 and ∞. To complete the specification of an element of the radiation it is necessary to indicate its plane of polarization. This will be determined if the intensities of the equivalent beams, polarized in any two mutually perpendicular planes containing the direction of propagation, are given. We may fix these planes by making one of them that which contains the normal to some arbitrary fixed plane as well as the axis of propagation. The second plane is the perpendicular plane which contains the direction of propagation. We shall distinguish the intensities of the two corresponding plane polarized beams by the suffixes 1 and 2 respectively. In the light of these explanations we may express the energy, incident on dS per unit time, whose frequency lies between ν and $\nu + d\nu$, and whose direction of propagation lies in a cone of solid angle $d\omega$ described about an axis making an angle θ with the normal to dS, in the form

$$\{i_1(\nu\omega) + i_2(\nu\omega)\} \cos\theta \, dS \, d\nu \, d\omega \quad \ldots\ldots\ldots\ldots\ldots(1).$$

We shall now consider what happens to the radiation which falls on any small material object placed inside the enclosure at constant temperature which contains the radiation whose properties we are investigating. Part of this radiation will be reflected from the surface and part may escape after penetrating the interior and undergoing refraction, internal reflection and so on, but the remainder will be absorbed. By absorption of radiation we understand its conversion into some non-radiant form of energy, so that it becomes temporarily stored in the matter. Let A denote the proportion of the incident energy which is not absorbed. In general A will depend on the nature of the substance, the frequency of the radiation, the plane of polarization and the azimuth θ. It might also depend on the incident intensity, although this is usually assumed, on experimental grounds, not to be the case. For the whole surface the net absorption of energy in unit time may evidently be written

$$\int_S \int_0^\infty \int_0^{2\pi} \{i_1(1 - A_1) + i_2(1 - A_2)\} \cos\theta \, dS d\nu \, d\omega \quad \ldots(2).$$

The radiant energy emitted from the whole surface may be written in the form

$$\int_S \int_0^\infty \int_0^{2\pi} (\epsilon_1 + \epsilon_2) \cos\theta \, dS d\nu \, d\omega \quad \ldots\ldots\ldots\ldots\ldots(3),$$

where ϵ_1 and ϵ_2 may depend on the nature of the substance, the frequency ν, the angle θ, and the plane of polarization as well as on the temperature. ϵ_1 and ϵ_2 must be independent of i_1 and i_2 provided the enclosure contains radiating matter other than the body S and provided the area of S is vanishingly small compared with that of the other matter.

The only way in which the body S can either gain or lose energy is by interchange of radiation, so that the net rate at which its energy increases is

$$\int_S \int_0^\infty \int_0^{2\pi} \{i_1(1-A_1) - \epsilon_1 + i_2(1-A_2) - \epsilon_2\} \cos\theta \, dS \, d\nu \, d\omega \quad \ldots(4).$$

The value of this integral must always be zero, otherwise the temperature of S would alter. This would contravene the second law of thermodynamics; since the difference of temperature thus established could be used to furnish available work, which would then be obtained from a source at a constant temperature.

This conclusion must be true whatever the shape, size and nature of the body S, whatever its position in the enclosure and whatever the nature, shape and size of the enclosure and the other matter contained in it may be; provided only that the enclosure is maintained at a constant temperature.

The fact that the integral (4) vanishes under all these conditions enables us to establish many important properties of the functions i, A and ϵ. In the first place the radiation must be isotropic; that is to say, i must have the same value at a given point for all directions in space. It cannot be a function of ω. If it were, the intensity of the radiation in some directions would be stronger than in others. By taking S to be a flat object it could be turned so as to receive more or less of the stronger radiation in different positions. Differences of temperature would thus be set up which would contravene the second law. This result must be true for all frequencies and all planes of polarization. This is clear since S may be a substance which absorbs some frequencies and transmits others, or it may be a plate of a material like tourmaline which has much more intense absorption for light polarized in certain planes than in others.

By expressing the element of solid angle $d\omega$ in terms of the angle θ between the direction of the ray and the normal to dS we may write (4) in the form

$$2\pi \int_S \int_0^\infty \int_0^1 \{i_1(1-A_1)-\epsilon_1+i_2(1-A_2)-\epsilon_2\}\, x\,dS\,d\nu\,dx = 0 \ldots (5).$$

By considering bodies of varying composition to make up S we can vary ϵ_1, ϵ_2, A_1 and A_2 independently of i_1 or i_2 as functions of either S, ν, x or the plane of polarization. It follows that the equation

$$i(1-A)=\epsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

is identically true for every wave-length, plane of polarization and element of angle. Thus for any assigned range of each of these quantities the part of the radiation incident on dS which is absorbed is equal to the similar radiation which is emitted from dS.

Next consider the case of a body which absorbs all the radiation which falls on it. Such a body has been called by Kirchhoff "perfectly black." The term black body is convenient, although it may be rather a misnomer, as such a body may be very bright when the illumination is due to its own temperature radiation.

For a black body, then, we have $A=0$, so that $i=\epsilon$. It follows that the intensity i of the radiation in the enclosure is equal to the emissivity ϵ, as defined by the preceding equations, of a perfectly black body. It is clear that ϵ must have the same value for all perfect absorbers at the same temperature; so that i must be independent of the nature of the materials present in the enclosure. Thus i is a function only of the frequency and plane of polarization of the radiation and *of the temperature T of the enclosure.* By symmetry $i_1=i_2$, so that the way in which the plane of polarization enters into i is a very simple matter. The determination of i as a function of ν and T will be considered later.

If α denotes the proportion of the incident radiation which is absorbed, then $\alpha = 1-A$ for each wave-length and so on. Whence $\epsilon/\alpha = i$ has the same value for all substances at the same temperature. Thus the emissive power divided by the absorption coefficient for any substance depends only on the frequency and plane of polarization of the radiation and the temperature, and is independent of the nature of the substance. This result, which was discovered by Balfour Stewart, is usually known as Kirchhoff's Law.

The foregoing results require amplification when the nature of the medium varies from one point of the enclosure to another. The relation $i = \epsilon$ no longer proves that i has the same value everywhere in the enclosure, since the emissivity ϵ of a black body may depend, and in fact does depend, on the nature of the surrounding medium. If we call I the function which has the same value at all points in the enclosure it is clear that the relation between i and I must be determined by properties of the medium which have nothing directly to do with absorption or emission. For if we consider a portion of the medium where there is no emitting or absorbing matter, but which is characterized by a particular value of the velocity of transmission, the appropriate value of i will somehow have to be established. In order to find I it is not therefore necessary to concern ourselves directly with the way in which ϵ may be modified according to the nature of the medium in which the emitting system is embedded. All that is necessary is to consider the passage of radiation across the interface between two portions of the medium characterized by different velocities of transmission.

Let us imagine that one of the regions in question is separated from the other by a perfectly reflecting interface. The introduction of this cannot make any difference to the nature of the radiation. The interface is punctured at the point A, leaving a small opening of area dS. In the upper medium, where the velocity of transmission is V, is a perfectly reflecting hemisphere with equal and symmetrical apertures at B and C which subtend equal infinitesimal solid angles $d\omega$ at A. If AD is the direction of the refracted ray corresponding to the incident ray BA, it is clear that with this arrangement the only radiation which can get from the lower medium to the part of the upper medium outside the hemisphere is that comprised in a small solid angle $d\omega'$ about AD and the equal beam which is symmetrical about the normal AF. The rest is all reflected by the hemisphere and returned through dS.

The ratio of the two elements of solid angle is

$$d\omega'/d\omega = \sin\theta' \, d\theta'/\sin\theta \, d\theta,$$

and the law of refraction gives

$$\sin\theta'/\sin\theta = V'/V.$$

If $i_\nu d\nu$ denotes the intensity of the incident radiation of frequency between ν and $\nu + d\nu$, in the upper medium, then the energy of this range of frequency which is incident on dS in unit time is $i_\nu \, d\nu \, dS \cos\theta \, d\omega$. The energy reflected along AC is

$$\rho i_\nu \, d\nu \, dS \cos\theta \, d\omega,$$

where ρ is the coefficient of reflection. If ρ' is the coefficient of reflection for the lower medium the proportion of incident energy which is refracted is $1 - \rho'$, so that the energy transmitted into the upper medium is

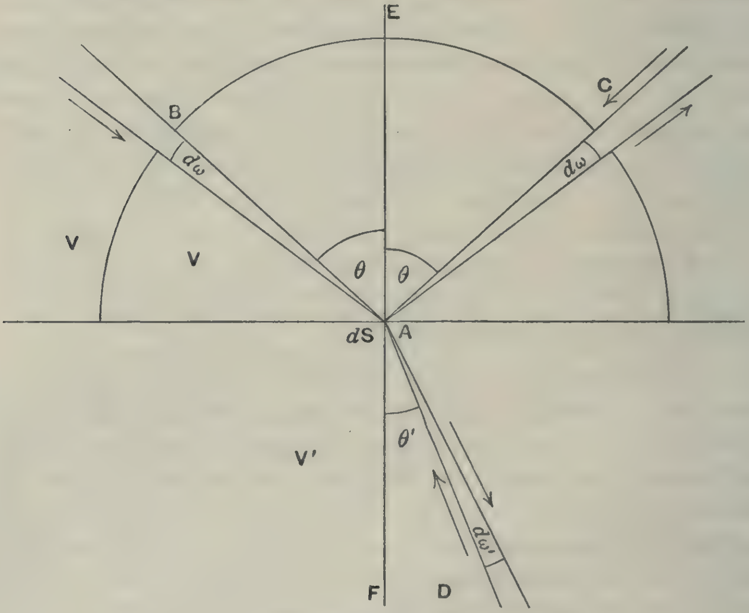


Fig. 40.

$$(1 - \rho') \, i_\nu' \, d\nu \, dS \cos\theta' \, d\omega',$$

where i_ν' is the value of i_ν for the lower medium. If the distribution of energy in the spectrum of the radiation in the two media is to remain invariable, we shall have

$$(1 - \rho) \, i_\nu \, d\nu \, dS \cos\theta \, d\omega = (1 - \rho') \, i_\nu' \, d\nu \, dS \cos\theta' \, d\omega'.$$

Thus
$$\frac{i_\nu \cos\theta\, d\omega}{i_\nu' \cos\theta'\, d\omega'} = \frac{1-\rho'}{1-\rho},$$

or
$$\frac{V^2 i_\nu}{V'^2 i_\nu'} = \frac{1-\rho'}{1-\rho} \qquad\qquad (7).$$

If the composition of the radiation in either medium were to change, the temperature of a material particle in it which possessed selective absorption would be altered. This would contravene the second law of thermodynamics, so that equation (7) must be true.

Now the left-hand side of (7) is independent of the direction and plane of polarization of the radiation, since the latter is isotropic at any given point. Thus if we can determine the value of the fraction on the right-hand side for some particular angle of incidence and plane of polarization we shall have determined it for all values. If θ is the polarizing angle and the light is polarized perpendicular to the plane of incidence no light is reflected at the boundary and $\rho = \rho' = 0$. It follows that ρ must always be equal to ρ' and $V^2 i_\nu$ must always be equal to $V'^2 i_\nu'$, so that the quantity which is invariable, which we previously denoted by I, is not the intensity i_ν but $i_\nu \times V^2$. Thus $V^2 i_\nu$ must be a universal function $F(\nu, T)$ of ν and T which we shall seek to determine. Also, if L_ν is the energy per unit volume of the radiation whose frequency lies between ν and $\nu + d\nu$, we observe that $i_\nu \propto V L_\nu$, so that $V^3 L_\nu$ is also a universal function of ν and T.

Stefan's Law.

We shall first find how the density of the complete radiation depends upon the absolute temperature T. Suppose that at some point of the wall of our enclosure at constant temperature T there is fitted a cylinder in which a piston may be made to work up and down. Both the walls of the cylinder and the face of the piston are perfect reflectors of radiation. The effect of an outward motion of the piston is to increase the volume which is filled by the radiation and, in addition, work is done by the pressure p of the latter on the piston head. If the increment of the entropy of the system which is produced by a small displacement is denoted by dS, we have

$$dS = \frac{dU + p\,dv}{T} \qquad\qquad (8),$$

where dU is the increment in the internal energy and dv is the increment in the volume. If L is the energy of the complete radiation per unit volume, then $dU = Ldv$. We have seen (Chap. x, p. 212) that the pressure of isotropic radiation on a surface bounding it is $p = \frac{1}{3}L$. Since the process contemplated is a reversible one dS must, from thermodynamic principles, be a perfect differential; so that

$$\frac{\partial^2 S}{\partial v \partial T} = \frac{\partial^2 S}{\partial T \partial v}.$$

But
$$\left(\frac{\partial S}{\partial v}\right)_T = \frac{1}{T}\left[\left(\frac{\partial U}{\partial v}\right)_T + p\right] = \frac{4}{3}\frac{L}{T},$$

and
$$\left(\frac{\partial S}{\partial T}\right)_v = \frac{1}{T}\left(\frac{\partial U}{\partial T}\right)_v,$$

so that
$$\frac{4}{3}\frac{\partial}{\partial T}\left(\frac{L}{T}\right) = \frac{\partial}{\partial v}\left[\frac{1}{T}\left(\frac{\partial U}{\partial T}\right)_v\right] = \frac{1}{T}\frac{\partial L}{\partial T},$$

and
$$\frac{\partial L}{\partial T} = 4\frac{L}{T} \quad \text{.................................(9)},$$

or
$$\log L = 4 \log T + \text{const.}$$

Thus
$$L = AT^4 \quad \text{.................................(10)}.$$

It follows that the energy per unit volume, *in vacuo*, of the radiation in equilibrium in an enclosure at the absolute temperature T, is equal to a universal constant A multiplied by the fourth power of the absolute temperature. Since the intensity of the radiation is equal to the energy per unit volume multiplied by the velocity of light, it follows that the former must also be proportional to the fourth power of the absolute temperature. Moreover, if E is the total emission from unit area of a perfectly black body, we see from p. 330 that $E = A'T^4$, where A' is a new universal constant. This result is usually known as Stefan's Law. It was suggested by Stefan*, in the inaccurate form that the total radiant energy emission from bodies varies as the fourth power of the absolute temperature, as a generalization from the results of experiments. The credit for showing that it is a consequence of the existence of radiation pressure combined with the principles of thermodynamics is due to Bartoli† and Boltzmann‡.

---

* *Wiener Ber.* vol. LXXIX. p. 391 (1879).

† Bartoli, *Sopra i movimenti prodotti dalla luce etc.*, Firenze, 1876.

‡ *Ann. der Phys.* vol. XXII. p. 31 (1884).

Since L, i and ϵ may be split up into their spectral components it follows that each of these will have to satisfy equations of the type

$$\int_0^\infty (L_\nu, i_\nu, \epsilon_\nu)\, d\nu = \text{const.} \times T^4 \ldots\ldots\ldots\ldots\ldots(11).$$

Reflection of Radiation at a Moving Mirror.

There is another effect, in addition to those already contemplated, which is produced when light or radiation is reflected at a moving surface. This results in a change in the wave-length or frequency of the light, which is akin to Doppler's effect. By considering the simple case in which the mirror is moving parallel to the direction of propagation of the radiation, which is incident normally, it is clear that the frequency of the light is diminished after reflection by an amount which is proportional to the velocity of the reflector, when this is in the same sense as that of the incident radiation. If the reflector is moving towards the incident beam the frequency of the reflected beam is greater than that of the incident beam. The complete resolution of more complex cases is effected most easily by means of the principle of relativity.

We shall consider the case of a plane reflector moving with uniform velocity v parallel to the direction of its normal, which we shall take as the axis of x. In the case of a ray incident at an angle θ the vectors which specify it will only contain the space and time coordinates through the factor

$$\cos n\left(t + \frac{x\cos\theta + z\sin\theta}{c} + \gamma\right),$$

if the plane of incidence is that of the x and z axes and γ is a constant which specifies the phase. The corresponding factor for the reflected ray is

$$\cos n\left(t - \frac{x\cos\theta - z\sin\theta}{c} + \gamma\right),$$

provided the reflector is at rest. If the reflector is in motion these expressions must be unchanged when all the quantities are measured in reference to axes x', y', z' and t' referred to which the mirror is at rest. The relation between x', y', z' and t' and x, y, z and t is given by the equations

$$x' = \beta(x - vt),\ y' = y,\ z' = z,\ t' = \beta(t - vx/c^2) \text{ and } \beta = (1 - v^2/c^2)^{-\frac{1}{2}}.$$

Thus referred to axes moving with the mirror, the vectors which specify an incident ray will be expressed by

$$a \cos n' \left(t' + \frac{x' \cos\theta' + z' \sin\theta'}{c} + \gamma'\right),$$

where a is a constant, the corresponding quantity for the reflected ray being

$$a \cos n' \left(t' - \frac{x' \cos\theta' - z' \sin\theta'}{c} + \gamma'\right).$$

In terms of the coordinates referred to axes at rest we shall have for the incident ray

$$a \cos n' \left\{\beta(t - vx/c^2) + \frac{\beta(x - vt)\cos\theta' + z\sin\theta'}{c} + \gamma'\right\}$$

$$= a \cos n' \left\{\beta\left(1 - \frac{v}{c}\cos\theta'\right)t + \frac{\beta(\cos\theta' - v/c^2)x + z\sin\theta'}{c} + \gamma'\right\},$$

and for the corresponding reflected ray

$$a \cos n' \left\{\beta(t - vx/c^2) - \frac{\beta(x - vt)\cos\theta' - z\sin\theta'}{c} + \gamma'\right\}$$

$$= a \cos n' \left\{\beta\left(1 + \frac{v}{c}\cos\theta'\right)t - \frac{\beta(\cos\theta' + v/c^2)x - z\sin\theta'}{c} + \gamma'\right\}.$$

The frequencies referred to axes x, y, z and t at rest in space are proportional to the coefficients of t. Thus if ν_1 is the frequency of the incident and ν_2 of the reflected ray as measured by an observer referred to whom the mirror is moving with velocity v, we have

$$\frac{\nu_2}{\nu_1} = \frac{1 + \frac{v}{c}\cos\theta'}{1 - \frac{v}{c}\cos\theta'};$$

but from equations (11), Chap. XIV,

$$\cos\theta' = \frac{\cos\theta - \frac{v}{c}}{1 - \frac{v}{c}\cos\theta},$$

thus
$$\frac{\nu_2}{\nu_1}=\frac{1-v^2/c^2}{1-2\frac{v}{c}\cos\theta+\frac{v^2}{c^2}}=1+2\frac{v}{c}\cos\theta \text{ ....(12)},$$

neglecting $(v/c)^2$ and higher powers.

The corresponding relation between the wave-lengths is

$$\frac{\lambda_2}{\lambda_1}=1-2\frac{v}{c}\cos\theta \text{ ....(13)}.$$

Wien's Law.

The foregoing considerations, coupled with thermodynamical principles, enable us to take another step forward towards the discovery of the function $F(\nu, T)$. This advance is due to Wien*. The argument resolves itself into two parts. In the first place, if we start with an enclosure containing only thermal radiation, such as we have seen to be characteristic of some temperature T, and then alter the nature of this radiation by means of a motion of some part of the perfectly reflecting boundary wall, we shall be able to show that the resulting radiation is invariably such as is characteristic of some other undetermined temperature T'. Having established this proposition, the second step consists in making use of it so as to find out as much as we can about the nature of the function $F(\nu, T)$.

Consider the cylinder with perfectly reflecting walls shown in Fig. 41. The ends are closed by plates of radiating matter maintained at the temperatures T_1 and T_2 respectively. T_2 is greater than T_1. The transverse partitions are perfectly reflecting and are provided with shutters D_1 and D_2 which can be opened or closed at will. The partitions can also be caused to slide along the cylinder. We now imagine the following processes to occur:—(1) D_2 is shut and D_1 open. Then C is filled with radiation characteristic of T_2 and A and B with radiation characteristic of T_1. When equilibrium has become established D_1 is closed. (2) The radiation in C is allowed to push the piston F and so compress the radiation in B until the pressure of the radiation in B is equal to that in C. On account of the Doppler effect the nature of the radiation in B will have become changed by the motion. Let us examine the consequence of supposing that the radiation now in B is not homogeneous with that in C. Since the total pressures in B and C are equal, the pressure, or density, for some frequencies will be greater in B and for others greater in C. (3) Place some selectively reflecting material over the opening in D_2, choosing it so that it transmits more of the rays whose density is greater in B than in C. Open the shutter an instant and then close it. The pressure in C is now greater than that in B. (4) Allow C to expand until the total densities again become equal. External work can be done by this expansion. (5) Open the door D_2 and allow the two radiations to mix. Leaving the door open push F back to its position after the displacement in (2). Then (6) close D_2 and push F back until the pressure in B is equal to that in A. In this step the work lost in the second step is exactly recovered.

---

* *Berl. Ber.* 9 Feb. 1893.

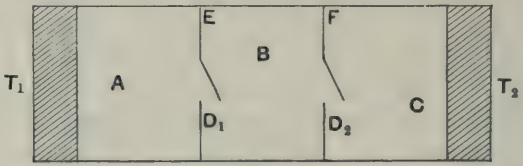


Fig. 41.

In the complete cycle, we have obtained work equal to that done by the expansion of the radiation after selective mixing without any transference of heat to the cold body T_1. The heat which is the equivalent of the work done must have come from the body T_2. As the process might be repeated indefinitely it is clear that the whole of the heat of T_2 might be converted into available work in this way. As this is contrary to the second law of thermodynamics it follows that the inequality supposed in (2) does not exist. Thus the following theorem must be true:—If we start with a perfectly reflecting enclosure containing nothing but radiation characteristic of a certain temperature, the nature of the radiation will be changed if the walls are allowed to expand or contract, but the resulting radiation will always be identical with that temperature radiation which exerts the same total pressure.

Now let us suppose that we have a cylinder fitted with a movable piston and filled with radiation characteristic of some temperature T. The face of the piston is perfectly smooth and perfectly reflecting. The walls of the cylinder reflect completely, i.e. without loss of energy, but irregularly in all directions. This device keeps the radiation isotropic and so makes it unnecessary for us to consider certain complications which would otherwise arise. The sectional area of the cylinder is A and its height h. h will vary with the position of the piston. Let $L(\lambda)\,d\lambda$ denote the energy per unit volume of the cylinder which belongs to wave-lengths between λ and $\lambda+d\lambda$, then $J=Ah\,L(\lambda)\,d\lambda$ is the total energy belonging to these wave-lengths which is present in the cylinder. Owing to the motion of the piston the value of J will tend to change. It will tend to decrease because the radiation whose wave-length is near λ will join some other group when it is reflected at the moving piston, and it will tend to increase because other radiation will have its radiation changed to values near λ when it is reflected. The rate of loss of energy by the group is clearly equal to the total amount which is incident on the moving piston in unit time. The proportion of J which belongs to rays whose direction of propagation lies within any small solid angle $d\omega=2\pi\sin\theta\,d\theta$ is $d\omega/4\pi$, since the radiation is isotropic. Thus the amount of energy belonging to these wave-lengths which is incident on the piston in unit time in directions lying within the cones whose semi-angles about the normal are θ and $\theta+d\theta$ is

$$\tfrac{1}{2}L(\lambda)\,d\lambda\,.\sin\theta\,d\theta\,.\,cA\cos\theta.$$

The rate of loss of energy by the group will be obtained by integrating this expression over all angles θ which lie between 0 and $\pi/2$. It is thus equal to

$$\tfrac{1}{4}Ac\,L(\lambda)\,d\lambda\,\text{........................}(14).$$

The calculation of the rate of gain of energy by the group is more complicated. If we consider a ray of wave-length λ' incident at an angle θ the wave-length of the reflected ray will be given by

$$\lambda'=\lambda\left(1-\frac{2v}{c}\cos\theta\right)\text{ by (13).}$$

The change of sign is due to the fact that we are now taking v to be positive for an outward motion of the piston, which corresponds to increasing volume. Thus, by reflection, rays characterized by λ', $d\lambda'$ and θ are changed into the group comprised between λ and $\lambda + d\lambda$, where

$$d\lambda' = d\lambda\left(1 - \frac{2v}{c}\cos\theta\right) \quad \text{...............(15).}$$

The energy incident in unit time in the group of waves defined by λ', $d\lambda'$, θ and $d\theta$ is

$$\tfrac{1}{2} Ac\, L(\lambda')\, d\lambda' \sin\theta \cos\theta\, d\theta \quad \text{...............(16).}$$

Only part of the energy of the λ' group is changed into energy of the λ group. The balance is equal to the work done by the group on the moving piston. The pressure exerted by the λ' group may be written

$$2Aj\cos^2\theta \text{ if } j = \frac{\sin\theta\, d\theta}{2} L(\lambda')\, d\lambda'.$$

The work done by this pressure per unit time is thus

$$2Aj\cos^2\theta \,.\, v.$$

By combining this result with (16) we see that the rate at which the energy of the group λ is increased by reflection from the group λ' is

$$Ajc\cos\theta - 2Aj\cos^2\theta\,.\,v \quad \text{..................(17).}$$

But $j = \frac{1}{2}\sin\theta\, d\theta\; L(\lambda')\, d\lambda'$ and

$$L(\lambda') = L\left(\lambda - \frac{2v}{c}\lambda\cos\theta\right), \text{ from (13),}$$

$$= L(\lambda) - \frac{2v}{c}\cos\theta\,.\,\lambda\frac{\partial L(\lambda)}{\partial\lambda} + \text{higher terms,}$$

by Taylor's theorem. By moving the piston slowly enough v/c can be made as small as we please, so that only the lower powers of this ratio need be considered. By substituting for $d\lambda'$ from (15) and neglecting higher powers of v/c than the first, we find

$$j = \frac{\sin\theta\, d\theta}{2}\left\{\left(1 - 2\frac{v}{c}\cos\theta\right)L(\lambda) - \frac{2v}{c}\cos\theta\,.\,\lambda\frac{\partial L(\lambda)}{\partial\lambda}\right\}d\lambda \quad \text{...(18).}$$

After substituting this value of j, (17) becomes

$$Ad\lambda\frac{\sin\theta\cos\theta\, d\theta}{2}\left\{\left(1 - 2\frac{v}{c}\cos\theta\right)L(\lambda) - \frac{2v}{c}\cos\theta\,.\,\lambda\frac{\partial L}{\partial\lambda}\right\}(c - 2v\cos\theta)$$

$$= \frac{Acd\lambda}{2}\sin\theta\cos\theta\, d\theta\left\{L(\lambda) - 2\frac{v}{c}\cos\theta\left(2L(\lambda) + \lambda\frac{\partial L(\lambda)}{\partial\lambda}\right)\right\}$$

to the same order in v/c. The total gain in energy is obtained by integrating this over all values of θ from 0 to $\pi/2$, giving

$$\frac{A}{4}d\lambda\left\{cL(\lambda)-\frac{4}{3}v\left(2L(\lambda)+\lambda\frac{\partial L(\lambda)}{\partial\lambda}\right)\right\} \quad \text{......(19).}$$

By combining this with (14) we see that the net increase of energy of the group λ in unit time due to an outward motion of the piston is

$$\frac{\partial J}{\partial t}=-\frac{A}{3}v\left\{2L(\lambda)+\lambda\frac{\partial L(\lambda)}{\partial\lambda}\right\}d\lambda \quad \text{.........(20).}$$

But by direct differentiation we have

$$\begin{aligned}\frac{\partial J}{\partial t}&=Ad\lambda\left\{L(\lambda)\frac{\partial h}{\partial t}+h\frac{\partial L(\lambda)}{\partial t}\right\}\\&=Ad\lambda\left\{vL(\lambda)+h\frac{\partial L(\lambda)}{\partial t}\right\}.\end{aligned}$$

Thus, from (20),

$$\frac{\partial L(\lambda)}{\partial t}=\kappa\left(-5L(\lambda)-\lambda\frac{\partial L(\lambda)}{\partial\lambda}\right) \quad \text{............(21),}$$

where $\kappa\equiv\dfrac{v}{3h}$ is a function only of the time t.

The differential equation (21) gives the relation between $L(\lambda)$ and the time t as the piston is moved. As we have seen, the effect of the motion is to change the radiation in such a way that it is always identical with the radiation which characterizes some temperature T. Thus for each time t there is a corresponding temperature T, and T must, therefore, be a function of t only. Thus it must be possible to replace t in (21) by some function of T only. This can be determined if we make use of Stefan's Law together with certain simple properties of the radiation which may be regarded as given by experiment.

Consider the total energy density $L=\int_0^\infty L(\lambda)\,d\lambda$. Integrating both sides of (21) with respect to $d\lambda$ from 0 to ∞, we have

$$\frac{\partial L}{\partial t}=-\kappa\left\{5L+\Big|\lambda L(\lambda)\Big|_0^\infty-L\right\}.$$

Now the form of the experimental curves connecting $L(\lambda)$ and λ shows that the product $\lambda L(\lambda)$ vanishes at both limits, so that

$$\frac{\partial L}{\partial t}=-4\kappa L.$$

But by Stefan's law $L = A_1 T^4$, where A_1 is a universal constant. Thus

$$4A_1 T^3 \frac{\partial T}{\partial t} = \frac{\partial L}{\partial t} = -4\kappa A_1 T^4$$

or

$$-\kappa\, dt = \frac{dT}{T};$$

so that (21) may be written

$$T\frac{\partial L(\lambda)}{\partial T} = 5L(\lambda) + \lambda\frac{\partial L(\lambda)}{\partial \lambda} \quad \dots\dots\dots\dots\dots(22).$$

If the variables are changed to $T_1 = \log T$, $\lambda_1 = \log \lambda$ and $L_1 = \log L(\lambda)$, (22) becomes

$$\frac{\partial L_1}{\partial T_1} = 5 + \frac{\partial L_1}{\partial \lambda_1}.$$

Now change the independent variables from T_1 and λ_1 to T_1 and π_1, where $\pi_1 = T_1 + \lambda_1$. Denoting the new partial derivatives by δ, we have

$$\frac{\partial L_1}{\partial \lambda_1} = \frac{\delta L_1}{\delta \pi_1}\frac{\delta \pi_1}{\delta \lambda_1} + \frac{\delta L_1}{\delta T_1}\frac{\delta T_1}{\delta \lambda_1} = \frac{\delta L_1}{\delta \pi_1},$$

and

$$\frac{\partial L_1}{\partial T_1} = \frac{\delta L_1}{\delta \pi_1}\frac{\delta \pi_1}{\delta T_1} + \frac{\delta L_1}{\delta T_1}\frac{\delta T_1}{\delta T_1} = \frac{\delta L_1}{\delta \pi_1} + \frac{\delta L_1}{\delta T_1},$$

so that

$$\frac{\delta L_1}{\delta T_1} = 5,$$

and

$$L_1 = 5T_1 + F(\pi_1),$$

where $F(\pi_1)$ is an arbitrary function of π_1. Substituting for the logarithms and multiplying both sides by $d\lambda$ we find that

$$\begin{aligned} L(\lambda)\, d\lambda &= T^5 \phi(T\lambda)\, d\lambda \dots\dots\dots\dots\dots\dots\dots(23) \\ &= \lambda^{-5}(T\lambda)^5 \phi(T\lambda)\, d\lambda \\ &= \lambda^{-5}\chi(T\lambda)\, d\lambda \quad \dots\dots\dots\dots\dots\dots(24), \end{aligned}$$

where ϕ and χ are undetermined functions of the product $T\lambda$ only.

It is known from experiment that $L(\lambda)$ has the value zero when $\lambda = 0$ or ∞ and a single maximum between these limits. Differentiating (24) by λ we see that the maximum and minimum values of $L(\lambda)$ satisfy the equation

$$\frac{\partial L(\lambda)}{\partial \lambda} = \lambda^{-6}\left\{\lambda T\frac{\partial[\chi(\lambda T)]}{\partial(\lambda T)} - 5\chi(\lambda T)\right\} = 0.$$

The value zero at $\lambda=\infty$ is clearly indicated, but the value at $\lambda=0$ cannot be foreseen without further information about the function χ. Presumably the maximum value is given by

$$\lambda T\frac{\partial[\chi(\lambda T)]}{\partial(\lambda T)}-5\chi(\lambda T)=0 \quad\ldots\ldots\ldots\ldots\ldots(25).$$

If this equation has a single root $\lambda T=b$, then the value λ_m of λ, for which $L(\lambda)$ is a maximum, will vary with the temperature in such a way that the product $\lambda_m T$ is always constant. This result, which has been well established by experiment, is known as Wien's displacement law.

It is probable that the relations (23) and (24) which involve the universal undetermined functions ϕ and χ of the argument λT are as far as we can get, from such very general considerations as have been employed above. To determine these functions more particularly it is probably necessary to consider the constitution of the radiating matter in a manner more explicit than we have done hitherto.

The Formula of Rayleigh and Jeans.

We have seen that the properties of the radiation in an enclosure at a uniform temperature are determined by the temperature alone and are independent of the nature of the material bodies which are present in the enclosure. Thus if the enclosure contains nothing but radiation, the latter will have exactly the same constitution at a given temperature whatever the walls are made of, provided they have even the smallest power of emitting every possible kind of radiation. On the other hand it is evident that the nature of the contained radiation, when equilibrium has been established, must be determined as the result of an equality between the emission from and absorption by the walls. The constitution of the radiation ought therefore to be determinate if we can calculate the rate of emission and absorption of different types of radiation in any particular case. Since the nature of the radiation is independent of that of the matter it makes no difference what constitution we assume for the matter, which we make use of in carrying out the calculations, provided that it is a possible type of matter and also is one which has some capacity for emission in every part of the spectrum. To facilitate the calculations one naturally assumes the simplest type of hypothetical matter which is compatible with these requirements.

Of the attempts to solve the radiation problem which have been based on the principle thus outlined, the earliest to rest on a substantial foundation, and the most successful, is due to Planck. Before considering Planck's theory we shall briefly indicate another method of attack which has led to results that are quite inconsistent with the experimental data. It is a well-known result in molecular dynamics (see Chap. XVII) that if any self-contained dynamical system possessing sufficient complexity is provided with a certain amount of energy and left for a sufficient length of time, a state of statistical equilibrium will finally become established. This state of equilibrium is characterized by the fact that each degree of freedom, or each coordinate which is required to specify completely the whole energy of the system, possesses the same average amount of energy. Now consider a perfectly reflecting enclosure containing a small amount of matter. The matter contains a finite number N, let us say, of molecules and each of these will have some finite number, p on the average, of degrees of freedom. On the other hand the aether which the enclosure contains will be capable of an infinite number of modes of vibration. These are determined by the geometry of the perfectly reflecting boundary and extend from the gravest mode of vibration to vibrations of infinite quickness. There is thus an infinite number of degrees of freedom in the aether, whereas the number in the matter remains finite. Consequently, since each degree of freedom receives equal energy, all the energy will be found in the aether, in the final equilibrium state, and none in the matter. Moreover, for a given range $d\lambda$, there are many more possible modes of vibration the smaller λ is, so that all the energy tends to accumulate in the waves of infinitesimal wave-length. By calculating the number of natural wave-lengths between λ and $\lambda + d\lambda$ and by supposing that there is an infinite amount of energy in the whole system, so that each wave-length acquires the amount RT which is appropriate to two degrees of freedom in the matter at the same temperature, we can find the amount of energy which occurs in the stretch of radiation between

λ and $\lambda + d\lambda$. By carrying out a calculation of this kind, Jeans* has shown that

$$L(\lambda)\,d\lambda = \frac{8\pi RT}{\lambda^4}\,d\lambda \quad \ldots\ldots\ldots\ldots\ldots\ldots(26),$$

or in terms of frequency instead of wave-length

$$L(\nu)\,d\nu = 8\pi\frac{\nu^2}{c^3}RT\,d\nu \quad \ldots\ldots\ldots\ldots\ldots\ldots(27).$$

A rather similar conclusion had previously been reached by Lord Rayleigh†.

From this point of view the radiation problem reduces to a determination of the number of modes of vibration in the aether comprised within given limits of frequency. A simplified form of Jeans's calculation which is due to H. A. Lorentz‡ may be stated in outline as follows: Consider the temperature radiation in equilibrium with a small amount of matter in a rectangular box, whose sides are parallel to the coordinate axes and of lengths d_1, d_2, d_3. The walls are smooth and perfectly conducting so that no radiation is absorbed or emitted by them. In the steady state the box will be filled with stationary waves, satisfying the condition that the tangential electric intensity vanishes at every point of the boundary. Any parallel beam of radiation which is travelling in any one of the eight directions given by the combinations of the direction cosines $\pm l$, $\pm m$, $\pm n$, will always after reflection be travelling in one or other of this group of directions. The totality of such groups will therefore represent the number of modes of vibration of the aetherial part of the system. On account of the two planes of polarization which are required fully to specify the beam of radiation, each mode will have associated with it an amount of energy $2RT$ equal to four times that of a single degree of freedom. If λ is the wave-length of the radiation under immediate consideration, the condition at the conducting boundary requires that

$$\frac{2ld_1}{\lambda} = k_1, \quad \frac{2md_2}{\lambda} = k_2, \quad \frac{2nd_3}{\lambda} = k_3,$$

---

* *Phil. Mag.* vol. x. p. 91 (1905).

† *Phil. Mag.* vol. XLIX. p. 539 (1900); *Coll. Papers*, vol. IV. p. 483.

‡ *Theory of Electrons*, p. 93.

where k_1, k_2, and k_3 are integers. Since $l^2 + m^2 + n^2 = 1$, we have

$$\frac{k_1^2}{d_1^2} + \frac{k_2^2}{d_2^2} + \frac{k_3^2}{d_3^2} = \frac{4}{\lambda^2}.$$

Thus k_1, k_2, and k_3 can have any integral values which satisfy this equation. Since the equation is that of an ellipsoid whose coordinates are k_1, k_2, and k_3, and whose semi-axes are $2d_1/\lambda$, $2d_2/\lambda$, and $2d_3/\lambda$, the number of vibrations whose wave-lengths lie between λ and $\lambda + d\lambda$ is equal to the number of points having coordinates which are positive integers which lie in the volume between the ellipsoids whose parameters are determined by λ and $\lambda + d\lambda$ respectively. The number of such points is equal to the volume lying in a single octant between the two ellipsoidal shells. The number is therefore $4\pi d_1 d_2 d_3 d\lambda/\lambda^4$ and the energy per unit volume of the box and wave-lengths between the limits λ and $\lambda + d\lambda$ is $8\pi RT d\lambda/\lambda^4$. It has been found that every purely dynamical method of calculating the radiation formula leads to (26) and (27).

Although the value of $L(\lambda)$ given by (26) satisfies Wien's functional relation (24) it is quite inconsistent with the results of experiments on the complete temperature radiation. The experiments show that $L(\lambda)$ has a maximum for a value of λ whose position is governed by Wien's displacement law and has the value zero when $\lambda = 0$. But (26) would make $L(\lambda)$ increase indefinitely as λ approached zero. There are a number of ways in which this contradiction might conceivably be avoided. Jeans has suggested that the final state of equilibrium, if it could ever be attained, would be given by (26); but that the rate of emission of the energy of short wave-lengths by matter is so slow that the distribution given by (26) would only become established after an indefinite lapse of time, and in any finite time would not even be approximately realized. According to this view the observed distribution is one of intermediate equilibrium due to less fundamental causes. Nevertheless the causes which establish the intermediate distribution must be sufficiently fundamental to be common to all types of matter; otherwise the degree of consistency which has been observed in experiments on thermal radiation would not be found.

Another way of escape is to deny the applicability of the theorem of equipartition of energy to aetherial, as opposed to material, systems. There are a number of reasons why this is not a very satisfactory alternative. In the first place the same kind of objection may be made against the applicability of equipartition to material systems. Thus the specific heats of gases are not what would be given by a naïve application of this theorem, particularly when one considers the large number of electrons present in the atoms and that each electron ought to have its quotum of energy. On the other hand the law of equipartition has been so successful in other directions that it is difficult to believe that its deduction from dynamical principles involves a fundamental error. In the second place the dilemma in question does not seem to be connected with the law of equipartition alone, so that denial of the validity of this would not really remove the difficulty. It appears that a number of calculations which use the first method, that of equilibrating the absorption and emission of energy, lead to Rayleigh's formula (see for example Chap. XVII, p. 433). In fact this formula appears inevitably to arise whenever the emission and absorption of radiant energy by matter is assumed to be a continuous process subject to dynamical and electrodynamic laws.

Although it may appear very revolutionary to some, it seems to the writer that the only logical way out of these difficulties is to deny the adequacy of dynamics and electrodynamics for the explanation of the emission or absorption of radiation by matter. If we leave the familiar guidance of dynamics behind it is necessary to find some other set of fundamental principles to rely on. Although the matter cannot yet be regarded as fully and satisfactorily worked out, a valuable start in this direction has unquestionably been made by Planck and his followers.

Planck's Law.

We have stated that Planck has shown how theoretically to deduce a radiation formula which is in good agreement with the results of experiment. It would take up too much space adequately to consider the various vicissitudes which this theory has undergone at the hands of Planck and his critics, so that we shall practically confine ourselves to the discussion of the most recent form* of it.

---

* *Ann. der Phys.* vol. XXXVII. p. 642 (1912).

Although this involves special assumptions which seem rather strange at first sight, it is free from self-contradiction and from assumptions, such as that of the discontinuous nature of energy, which appear to do violence to the fundamental ideas of physics. So much could not be said of the earlier forms.

We shall suppose the radiating properties of the matter to arise from the presence in it of an indefinite number of minute electrical oscillators. To fix our ideas we may take the picture afforded by the electron theory and suppose each oscillator to consist of an electron in equilibrium in a certain position, when free from the action of external forces, and subject, when displaced, to a restoring force which is proportional to the displacement from the equilibrium position. If s denotes the moment of the doublet which is equivalent to the displaced electron, then the potential and kinetic energies are equal respectively to $\frac{1}{2}Ms^2$ and $\frac{1}{2}N\dot{s}^2$, where M and N are constants. The total energy is

$$U = \tfrac{1}{2}Ms^2 + \tfrac{1}{2}N\dot{s}^2 \quad\ldots\ldots\ldots\ldots(28).$$

If the component, along the axis s of the doublet, of the external force due to the incident radiation is E_s, the differential equation satisfied by s is

$$Ms + N\ddot{s} = E_s \quad\ldots\ldots\ldots\ldots(29),$$

if we neglect the reaction arising from the radiation. The natural frequency ν_0 of the oscillator is

$$\nu_0 = \frac{1}{2\pi}\sqrt{\frac{M}{N}} \quad\ldots\ldots\ldots\ldots(30).$$

When the small force arising from the reaction of the emitted radiation is included (see Chap. XII, p. 266) the equation for s becomes

$$Ms + N\ddot{s} - \frac{1}{6\pi c^3}\dddot{s} = E_s \quad\ldots\ldots\ldots\ldots(31).$$

The first step in the solution of our problem is to find the rate of emission and absorption of radiation by any particular oscillator.

The method originally adopted by Planck is an immediate application of the principles of electrodynamics which we have already developed. The radiation is determined by the acceleration of the electron, the rate of emission at any instant being proportional to the square of $\ddot{s}$ in accordance with formula (11) of

Chap. XII. The energy absorbed is equal to the work done by the external force E_s on the oscillator, so that the instantaneous rate of absorption of energy is $E_s \dot{s}$. Now the force E_s is the component which lies along the axis of the doublet of the electric intensity in the radiation. However complicated the radiation may be it will be possible to express E_s as a function of the time t by means of a Fourier's series in the form

$$E_s = \sum_{n=1}^{\infty} C_n \cos\left(\frac{2\pi nt}{T} - \delta_n\right) \text{..............} (32),$$

where the C's and δ's are undetermined constants and T is an arbitrary time which is greater than every t. By substituting this value of E_s in (31) and solving for s, we can show* that the mean value U of the energy of the resonator, which is equal to $M\dot{s}^2$ since the kinetic and potential energies must be equal on the average, is

$$U = \frac{3c^3}{16\pi\nu_0^2} T\overline{C_{n0}^2} \text{ ....} (33),$$

where $\overline{C_{n0}^2}$ denotes the mean value of all the coefficients C_n for which n lies near $\nu_0 T$. It happens that the distant terms do not contribute anything to the energy. But the energy $\int_0^\infty L(\nu)\,d\nu$ per unit volume of the radiation when expressed in terms of the electric and magnetic intensities is equal to half the mean value of $E_x^2 + E_y^2 + E_z^2 + H_x^2 + H_y^2 + H_z^2$. Since the radiation is isotropic the mean value of each of these terms is the same, so that

$$\int_0^\infty L(\nu)\,d\nu = 3\overline{E_s^2} \text{ ....} (34).$$

By putting $d\nu = \Delta n/T$, where Δn is a large number, the integral on the left can be expressed as an infinite series, and by comparing with the value of $\overline{E_s^2}$ which results from (32), it follows that

$$L(\nu_0) = \tfrac{3}{2} T\overline{C_{n0}^2},$$

whence from (33),

$$L(\nu_0) = 8\pi \frac{\nu_0^2}{c^3} U \text{ ....} (35).$$

This result is in formal agreement with equation (27), since the oscillator has two degrees of freedom and the energy per degree of freedom when equipartition holds is $\frac{1}{2}RT$. Thus it appears that if equipartition of energy is established among the oscillators by actions going on in the matter (these may be supposed to be independent of radiation) we are again led to Rayleigh's formula. This formula in fact appears to result from every method of calculation which makes both the absorption and emission of the energy by the matter take place in a continuous manner. It is well to point out that the Rayleigh formula expresses the experimental results satisfactorily when ν is small and T is large, that is to say when the energy of the radiators is large. There is therefore at least an element of truth in it.

---

* Cf. Planck, *Wärmestrahlung*, pp. 118 *et seq.*, Leipzig (1906).

In order to arrive at a formula which does not make $L(\nu)$ infinite when ν is infinite, it is necessary to introduce discontinuity somewhere, and thus bring probability and entropy considerations to bear on the state of the radiant energy. In his earlier papers the assumptions made were equivalent to postulating that the energy itself had a discontinuous structure, but Planck has now shown that equivalent results may be obtained by merely supposing that the radiant energy is emitted by jumps, the absorption taking place continuously. As the emission of radiant energy might be expected to be conditioned by the breaking up of some structure present in the matter, this seems a very natural hypothesis.

The investigation which follows involves a number of additional hypotheses. In the first place we assume that the energy U of the oscillators is determined entirely by interchange of radiation: that the influence of any direct dynamical interaction of the oscillators, if it occurs, has negligible consequences. We shall assume that the absorption of energy by the resonators is continuous and follows the requirements of the classical laws of dynamics and electrodynamics. It is thus determined by the conditions laid down on p. 349. On the other hand we shall assume that emission of energy is not a continuous process but one which never takes place except when the energy U of an oscillator is an integral multiple of a certain element of energy $\epsilon = h\nu_0$. In this expression ν_0 is the frequency of the oscillator and h is a universal constant. The element of energy is thus proportional to the frequency of the oscillator. We also assume that when emission takes place the oscillator suddenly loses *all* of its energy. It is clear that the oscillator must not invariably emit when its energy is an integral multiple of ϵ, otherwise the integral multiple would never exceed unity. We shall assume that whenever one of the critical values is reached, the ratio of the probability that no emission takes place to the probability that emission takes place is proportional to the density $L(\nu_0)$ of the radiation surrounding the oscillator. If η is the probability that emission takes place, the probability that emission does not take place is $1-\eta$; so that

$$\frac{1-\eta}{\eta} = pL(\nu_0) \qquad \text{......................(36)},$$

where p is a constant quantity which we shall determine later. The mean energy U of an oscillator in the steady state may now be determined in terms of η, as follows:

Out of N oscillators selected at random that have completed an emission $N\eta$ will emit when they have accumulated a single element of energy, $N(1-\eta)\eta$ when they have accumulated two elements of energy, $N(1-\eta)^{n-1}\eta$ when they have accumulated n elements of energy, and so on. Thus, in the steady state, out of N oscillators selected at random simultaneously,

$$\begin{array}{rlccc} N\eta = NP_0 & \text{will possess energy between} & 0 & \text{and} & \epsilon, \\ N(1-\eta)\eta = NP_1 & \quad ,, \qquad ,, \qquad ,, \qquad ,, & \epsilon & ,, & 2\epsilon, \\ \cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots & \cdots & \cdots \\ N(1-\eta)^{n-1}\eta = NP_{n-1} & \quad ,, \qquad ,, \qquad ,, \qquad ,, & (n-1)\epsilon & ,, & n\epsilon, \end{array}$$

where $P_n = (1-\eta)^n\eta$ is the probability that the energy of an oscillator lies between $n\epsilon$ and $(n+1)\epsilon$. The mean energy of an oscillator is therefore

$$U = \sum_{n=0}^{\infty} P_n\left(n+\frac{1}{2}\right)\epsilon = \left(\frac{1}{\eta}-\frac{1}{2}\right)\epsilon \qquad \text{...........(37)},$$

since the average value of the fractions of an element of energy which intervene between any two consecutive integral multiples is $\frac{1}{2}\epsilon$. Thus from (36)

$$U = \{pL(\nu_0)+\tfrac{1}{2}\}\epsilon \qquad \text{....................(38)}.$$

We shall determine p so that (38) agrees with the corresponding equation (35) of the former theory when U is large. In this case the $\frac{1}{2}$ in (38) can be neglected, so that, by comparing with (35),

$$p = \frac{c^3}{8\pi\nu_0^2\epsilon} = \frac{c^3}{8\pi h\nu_0^3} \qquad \text{...................(39)}.$$

Thus the mean energy U of the oscillators, and also, since

$$\frac{1}{\eta} = 1 + pL(\nu_0),$$

the division of the energy among them in the steady state, is completely determined.

In order to introduce the temperature T it is necessary to calculate the entropy S of the system. This is equal (see Chap. XVII, pp. 400 and 407) to R times the logarithm of the probability of the system, defined in a similar way to that introduced by Boltzmann* into the kinetic theory of gases. The present case is a little different from that contemplated in the kinetic theory, inasmuch as it is only the part of the energy of the resonators which is an integral multiple of ϵ that is a matter of chance and therefore subject to probability considerations†. The probability sought is the number of ways in which the N resonators can be arranged so as to have the given distribution of energy units, subject to the condition that the same distribution arises whenever the resonators which have a given number of units are interchanged among themselves. This probability may be calculated as follows:

Along the positive energy axis lay off marks at the points $0, 1\epsilon, 2\epsilon \ldots m\epsilon \ldots \infty\,\epsilon$. From these marks draw lines perpendicular to the axis and on them lay off $NP_0, NP_1, NP_2 \ldots NP_m \ldots NP_\infty$ equidistant dots. Altogether there are $\Sigma NP_m = N$ dots and their distribution over the diagram is a geometrical representation of the way in which the resonators are distributed about the energy of the system. Since the dots which have equal numbers of units are considered to give rise to systems which are indistinguishable, the number of independent ways in which the N dots can be arranged to form the given distribution is

$$W = N! \div \Pi (NP_m)!$$

where Π denotes the continued product for all the values of m from 0 to ∞. Thus

$$S = R \log W = R \log N! - R\Sigma \log (NP_m)!,$$

---

* *Vorlesungen über Gastheorie*, p. 41, Leipzig (1896).

† Planck, *Sitzber. der K. Pr. Akad. der Wiss.* vol. XXXV. p. 723 (1911).

and since we may use Stirling's approximation for the factorials when N is large

$$S = R \log N! - R\Sigma \{NP_m (\log NP_m - 1) + \tfrac{1}{2} \log 2\pi NP_m\}$$
$$= R \{\log N! - N \log N + N - \tfrac{1}{2}\Sigma \log 2\pi N\} - NR\Sigma P_n \log P_n,$$

using $\Sigma P_n = 1$ and remembering that the neglected term

$$\frac{R}{2} \Sigma \log P_n$$

vanishes compared with the others when N is indefinitely large. The first term on the right depends only on N and is therefore constant. It may be assimilated with the undetermined constant which is always included in the entropy. The value of this constant does not enter into our calculations, so that, leaving it out of account, we have for the entropy S_N of the N resonators

$$S_N = -NR \sum_{m=0}^{\infty} P_m \log P_m$$
$$= -RN \left\{\frac{1}{\eta} \log \eta + \left(\frac{1}{\eta} - 1\right) \log \left(\frac{1}{\eta} - 1\right)\right\}$$
$$= RN \left\{\left(\frac{U}{\epsilon} + \frac{1}{2}\right) \log \left(\frac{U}{\epsilon} + \frac{1}{2}\right) - \left(\frac{U}{\epsilon} - \frac{1}{2}\right) \log \left(\frac{U}{\epsilon} - \frac{1}{2}\right)\right\} \quad \text{...(40)}$$
$$= NS.$$

And since, by the definition of entropy,

$$\frac{1}{T} = \frac{dS}{dU} = \frac{R}{\epsilon} \log \frac{\dfrac{U}{\epsilon} + \dfrac{1}{2}}{\dfrac{U}{\epsilon} - \dfrac{1}{2}} \quad \text{..................(41),}$$

and $\epsilon = h\nu$,

$$U = \frac{h\nu}{2} \frac{e^{\frac{h\nu}{RT}} + 1}{e^{\frac{h\nu}{RT}} - 1} \quad \text{.....................(42).}$$

Finally, from (38) and (39),

$$L(\nu)\, d\nu = \frac{8\pi\nu^2 \epsilon}{c^3} \left(\frac{U}{\epsilon} - \frac{1}{2}\right) d\nu \quad \text{...............(43)}$$
$$= \frac{8\pi}{c^3} \frac{h\nu^3}{e^{\frac{h\nu}{RT}} - 1} d\nu \text{.....................(44),}$$

and

$$L(\lambda)\, d\lambda = \frac{8\pi ch\lambda^{-5}}{e^{\frac{ch}{RT\lambda}} - 1} d\lambda \quad \text{.....................(45).}$$

If $i_\nu d\nu$ is the intensity of a plane polarized constituent of the radiation, travelling in a given direction, we have $\frac{8\pi}{c} i_\nu = L(\nu)$, since the radiation is isotropic. Thus

$$i_\nu = \frac{c}{8\pi} L(\nu) = \frac{1}{c^2} \frac{h\nu^3}{e^{\frac{h\nu}{RT}} - 1} \quad \text{..............(46)}.$$

If the medium is one in which the velocity V of radiation differs from the value c which it possesses in a vacuum the right-hand side of (46) will have to be multiplied by c^2/V^2, since, according to the conclusions on p. 333, $V^2 i_\nu$ is the function which has a universal value. Corresponding changes would have to be introduced in (44) and (45). We shall, however, confine our discussion to the case of the radiation as it is found in a vacuum.

Formulae (44) and (45) are those which are known as Planck's radiation formulae. They are in agreement with the functional relations (23) and (24) demanded by Wien's argument. These formulae have been derived theoretically in other ways* which differ in important points from that which has just been given. But in order to obtain them it has always been found necessary to introduce discontinuity somewhere, either in the constitution of the radiation itself or in the mode of its absorption or emission.

When λT is small, or ν/T large, it is evident that unity may be neglected compared with the exponential in the denominator of the fraction, so that for small wave-lengths, or high frequencies, and low temperatures,

$$L(\nu)\, d\nu = 8\pi h \frac{\nu^3}{c^3} e^{-\frac{h\nu}{RT}} d\nu \quad \text{..............(47)},$$

$$L(\lambda)\, d\lambda = 8\pi \frac{ch}{\lambda^5} e^{-\frac{ch}{RT\lambda}} d\lambda \quad \text{..............(48)}.$$

When ν is large the exponential factor in (47) diminishes much more rapidly than ν^3 increases, so that $L(\nu) = 0$ when $\nu = \infty$. For the same reason $L(\lambda) = 0$ when $\lambda = 0$. Thus the conditions referred to on p. 341 are satisfied by Planck's formula. On the other hand when ν/T is small or λT large, (44) and (45) degenerate into

---

* For example cf. A. Einstein, *Ann. der Phys.* vol. XXII. p. 183 (1907).

$$\left.\begin{aligned} L(\nu)\,d\nu &= \frac{8\pi}{c^3}\,\nu^2 RT\,d\nu \\ L(\lambda)\,d\lambda &= 8\pi\,\frac{RT}{\lambda^4}\,d\lambda \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (49).$$

They are then identical with Rayleigh's formulae (26) and (27), which, as we have already pointed out, are in agreement with the results of experiments which are made subject to these conditions.

In fact throughout the whole range of the variables λ (or ν) and T, which has been tested, the differences between the experimental results and those given by Planck's formulae lie within the limits of experimental error.

Numerical Values of the Constants.

For the density in space of the black body radiation in a vacuum we have

$$L = \int_0^\infty L(\nu)\,d\nu = \frac{8\pi}{c^3}\int_0^\infty \frac{h\nu^3}{e^{\frac{h\nu}{RT}} - 1}\,d\nu$$

$$= \frac{48\pi R^4}{c^3h^3}\,\alpha T^4 = aT^4 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (50),$$

where
$$\alpha = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \ldots = 1{\cdot}0823.$$

The absolute constant a which occurs in the expression of Stefan's law is thus defined in terms of R, c and h. Accurate measurements of a have been made by Kurlbaum*, who finds†

$$a = 7{\cdot}06 \times 10^{-15}\ \text{erg cm.}^{-3}\ \text{deg.}^{-4}.$$

Another relation between the constants is given by the value of λ for which $L(\lambda)$ is a maximum. If we differentiate $L(\lambda)$ by λ and equate to zero, the equation which corresponds to (25) is

$$5\,(e^{-x} - 1) + x = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (51),$$

---

* *Ann. der Phys.* vol. LXV. p. 759 (1898).

† Some recent values of a are much higher than this. Thus Gerlach (1912) finds a value about 10 °/₀ higher, and Féry and Drecq (1911) find a value more than 20 °/₀ higher than Kurlbaum's. Most of the determinations, however, have given values of a which are not far from $7{\cdot}1 \times 10^{-15}$.

where $x = \frac{ch}{R\lambda T}$. (51) has two real roots, viz. $x = 0$ and $x = 4{\cdot}9651$. Thus the maximum value is given by substituting $x = 4{\cdot}9651$ in

$$\lambda_m T = \frac{ch}{Rx} = b \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(52).$$

According to the measurements of Lummer and Pringsheim* $b = 0{\cdot}294$ cm. deg. Solving (50) and (52) for R and h we find

$$\left.\begin{aligned} R &= 1{\cdot}346 \times 10^{-16} \text{ erg deg.}^{-1} \\ h &= 6{\cdot}548 \times 10^{-27} \text{ erg sec.} \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots(53).$$

Since R is the constant in the gas equation $pv = R_1 T$ reckoned for a single molecule, this value of R enables us to deduce the value of N, the number of molecules in a cubic centimetre of a gas under standard conditions, and the charge e on an ion, from well-known data. We have $pv = NRT$, where

$$p = 76 \times 13{\cdot}6 \times 981 \text{ dynes cm.}^{-2}, \quad v = 1 \text{ cm.}^3 \quad \text{and} \quad T = 273 \text{ deg.}$$

Whence $$N = 2{\cdot}76 \times 10^{19} \text{ per cm.}^3.$$

Since the charge which is required to liberate half a cubic centimetre of hydrogen, measured under standard conditions, in electrolysis is 0·4327 E.M. units (see Chap. I, p. 6) it follows that

$$Ne = 0{\cdot}4327 \text{ E.M.U.},$$

whence $$e = 4{\cdot}69 \times 10^{-10} \text{ electrostatic units}$$
$$= 18{\cdot}9 \times 10^{-10} \text{ in our units.}$$

These values of N and e are in excellent agreement with those which have been found by Millikan, Rutherford and others, using more direct methods. This agreement must be regarded as supporting very strongly those assumptions in Planck's demonstration which are necessary to produce the formulae finally obtained.

We shall see that Planck's radiation theory has recently received unexpected support in two other directions. One of these will now be considered.

Radiation and Specific Heat.

From the phenomena exhibited by absorption bands, the residual rays, and so on, we know that something like Planck's oscillators must exist in actual matter and possess natural frequencies not far from the frequencies of the radiation forming the visible spectrum. From (42) we can calculate the energy corresponding to any one of these natural frequencies at a given temperature.

---

* *Verhand. der Deutsch. Physik. Ges.* vol. I. p. 230.

Einstein* suggested that practically all the energy stored in simple bodies might belong to a few frequencies, and, on this hypothesis, was able to calculate the specific heat as a function of the frequency. If this hypothesis were not true it would be necessary to suppose that the number of systems having a given frequency is much smaller than the number of atoms present in the substance; otherwise the specific heats of bodies would be much larger than those which they actually possess.

Let us suppose that there are N oscillators per unit mass of any substance, and that they all have the same natural frequency ν. The total energy of the oscillators in unit mass of the substance at the temperature T is, by (42),

$$N\left(\frac{h\nu}{e^{\frac{h\nu}{RT}}-1}+\frac{h\nu}{2}\right).$$

If all the heat is in this kind of energy, the specific heat C_v will be given by differentiating this expression with respect to the temperature, so that

$$C_v = N\frac{\partial}{\partial T}\left(\frac{h\nu}{e^{\frac{h\nu}{RT}}-1}+\frac{h\nu}{2}\right) = N\frac{h^2\nu^2}{RT^2}\frac{e^{\frac{h\nu}{RT}}}{\left(e^{\frac{h\nu}{RT}}-1\right)^2} \quad \ldots(54).$$

At low temperatures, according to this formula,

$$C_v = N\frac{h^2\nu^2}{RT^2}e^{-\frac{h\nu}{RT}},$$

since we can neglect unity compared with the exponential in the denominator. Thus C_v rapidly becomes extremely small with decrease of temperature. At high temperatures C_v approximates to $NRe^{\frac{h\nu}{RT}} = NR$ approx. It thus becomes independent of the temperature; and if we suppose that there is one natural frequency per atom, so that N is the number of atoms per unit mass of the substance, the value NR agrees quantitatively with Dulong and Petit's Law.

---

* *Ann. der Phys.* vol. XXII. p. 180 (1907).

An elaborate investigation of the specific heats of a large number of substances, particularly at low temperatures, has recently been carried out by Nernst* and his pupils. They find that the relation between C_v and T is of the same general character as that called for by (54), but that a better agreement is obtained when a slightly different equation is used, viz.

$$C_v = \frac{NR}{2}\left\{\frac{\left(\frac{h\nu}{RT}\right)^2 e^{\frac{h\nu}{RT}}}{\left(e^{\frac{h\nu}{RT}}-1\right)^2} + \frac{\left(\frac{h\nu}{2RT}\right)^2 e^{\frac{h\nu}{2RT}}}{\left(e^{\frac{h\nu}{2RT}}-1\right)^2}\right\} \quad \text{.........(55).}$$

The second term, which was introduced by Nernst and Lindemann, was obtained empirically.

The agreement between the results of experiments and (55) is shown by the following numbers, which represent typical cases selected at random from a paper by Nernst and Lindemann†.

Copper (Cu)			NaCl			Diamond (C)		
Abs. Temp.	C_p (cal.)	C_p (obs.)	Abs. Temp.	C_p (cal.)	C_p (obs.)	Abs. Temp.	C_p (cal.)	C_p (obs.)
23·5	0·15	0·22	25·0	0·32	0·29	30	0·000	0·000
27·7	0·31	0·33	25·5	0·34	0·31	42	0·000	0·000
33·4	0·59	0·54	28·0	0·48	0·40	88	0·006	0·03
87·0	3·37	3·33	67·5	2·88	3·06	92	0·009	0·03
88·0	3·39	3·38	69·0	2·95	3·13	205	0·62	0·62
137	4·65	4·57	81·4	3·49	3·54	222	0·78	0·76
234	5·52	5·59	83·4	3·64	3·75	262	1·16	1·14
290	5·75	5·79	138	4·90	4·87	306	1·59	1·58
323	5·81	5·90	235	5·73	5·76	358	2·08	2·12
450	6·03	6·09	—	—	—	413	2·55	2·66
—	—	—	—	—	—	1169	5·41	5·45

C_p is the specific heat of the solid substance at constant pressure and is the quantity which is given by the experiments. The calculated values of C_p are obtained from the values of C_v given by (55), together with the known relation, based on thermodynamic considerations,

$$C_p = C_v\left(1 + 9\,\frac{\alpha^2 V}{KC_v}\,T\right) \quad \text{..................(56),}$$

---

* *Zeits. für Elektrochemie*, vol. XVII. pp. 265, 817 (1911).

† *Ibid.* p. 817.

where α is the coefficient of linear expansion, V is the atomic volume, and K the compressibility. The difference between C_p and C_v is only appreciable at the higher temperatures.

It will be observed that (55) contains only one adjustable constant, the frequency ν, so that the agreement shown by the table is quite convincing. In fact in the case of KCl, KBr and NaCl the value of ν has been taken from the experiments of Rubens and Hollnagel on the residual rays from these substances. Thus in these cases the complete thermal behaviour may be predicted from the determination of a single optical constant, and shows an excellent agreement with the observed results. In other cases, for instance mercurous chloride, it is necessary to use a formula involving a summation over two different values of ν; but there are good reasons for assigning at least two frequencies for compound substances, so that this cannot be considered to be an argument against the general position.

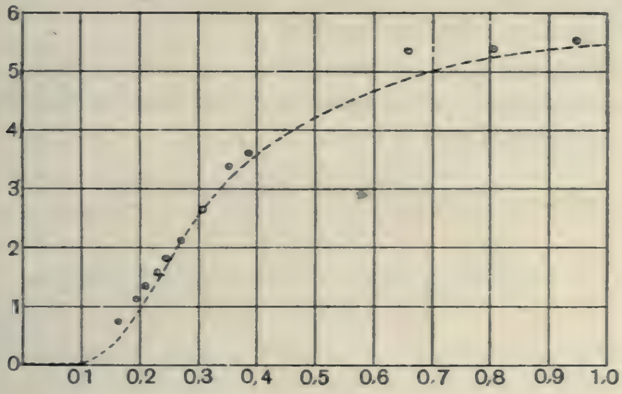


Fig. 42.

The general character of the relation between C_v and $\frac{h\nu}{RT}$, given by (54) and (55), is shown by the accompanying diagram (Fig. 42) taken from Einstein's paper. The ordinates represent the values of the right-hand side of (54) and the abscissae those of $\frac{h\nu}{RT}$. NR is taken $= 5{\cdot}94$. The calculated values of C_v are shown by the broken curve. The circles represent the older observations of the specific heat of carbon, an appropriate value of ν being assumed. The newer observations agree better with (55) than with (54), but the general character of the curves is the same.

In a recent paper P. Debye[*] has deduced formulae for the specific heats at various temperatures which agree even better than (55) with the experimental results. Debye's method, which is a very interesting one, brings the specific heats quite directly into relation with the elastic constants of the substances concerned. The formulae obtained are also quite simple and general, and only contain the same number of parameters as those of Einstein and Nernst-Lindemann. This also represents an important advance, as (54) is known to be inaccurate and (55) has never received a satisfactory theoretical explanation.

Another direction in which Planck's theories have received interesting support will be considered under the heading of photo-electric action in Chapter XVIII.

---

[*] *Ann. der Phys.* vol. XXXIX. p. 789 (1912).

CHAPTER XVI

THE THEORY OF MAGNETISM

When we regard magnetic phenomena from the point of view of the disturbance produced by the material media in which the effects take place we are struck by the great variety of phenomena manifested as compared with those in the electrostatic case. When a plane slab of dielectric is placed in a uniform electric field in free space, so that the lines of force are perpendicular to the face of the slab, the electric intensity is invariably smaller inside the dielectric than in the surrounding space. In dealing with magnetic phenomena an effect of this nature is by no means the invariable rule.

In comparing the behaviour of magnetic media they are found to belong to one of three distinct classes.

1. Diamagnetic media. Substances of this class are characterized by a permeability which is less than the value unity attributed to free space. They therefore tend to move into the weakest parts of the magnetic field, as this arrangement makes the potential energy of the system a minimum.

2. Paramagnetic media. The permeability is constant and greater than unity. These substances tend to move into the strongest parts of the field. Their behaviour is thus analogous to that of homogeneous dielectrics in the electrostatic case.

3. Ferromagnetic media. The permeability is greater than unity but is not constant. The polarization (magnetization) tends to reach a saturation value as the magnetic intensity is increased. The high value which the polarization may reach is a characteristic feature of this class of bodies.

There is another very important and fundamental difference in the magnetic and electric behaviour of matter. It is never possible to separate the positive and negative magnetic charges in different portions of matter in the way in which the electric charges can be separated. This would be extremely unlikely to be the case unless the fundamental magnetic element contained both positive and negative magnetism and were therefore similar to an electric doublet rather than to the positive and negative charges of which such a doublet is constituted.

This basic difference receives a ready explanation on the electron theory. According to that theory magnetic forces can arise only from the motion of electric charges and, in the last analysis, from the motion of electrons. Now there is no possible motion of electrons which can give rise to an isolated magnetic pole; but there is a very simple type of motion which gives rise to a system which is the magnetic analogue of the electric doublet, i.e. a system which has positive magnetism on one side of it and negative on the other. Many years ago Ampère built up a theory of magnetic media on the assumption that the atoms were the seat of circular electric currents. As is well known, such a current behaves like a small magnet, and the hypothesis is therefore all that is required to account for magnetic polarization and hence, from the analogy with dielectrics, for magnetizable media. Now we shall be able to show that an electron revolving in a closed orbit is equivalent to a small magnet in the same way that Ampère's atomic currents were. This theory, whose development is due largely to Weber and Langevin, will be shown to give a simple explanation of diamagnetism as well as paramagnetism. With certain further assumptions which do not seem improbable, it can be made to give a good account of the more complex phenomena of ferromagnetism as well.

The Magnetic Force due to a Moving Electron.

We have seen (Chap. XI, p. 221) that the components of the magnetic intensity due to an electrically charged particle moving with the uniform velocity w parallel to the z axis are

$$H_x = \frac{\beta}{\sqrt{1-\beta^2}}\frac{\partial\phi_1}{\partial y_1}, \quad H_y = -\frac{\beta}{\sqrt{1-\beta^2}}\frac{\partial\phi_1}{\partial x_1}, \quad H_z = 0,$$

where $\beta = w/c$. When β is small, which we shall suppose to be the case in the motions we are dealing with, we can put

$$\phi_1 = e/r, \quad x_1 = x \quad \text{and} \quad y_1 = y.$$

Thus

$$H_x = -\frac{w}{c}\frac{e}{r^2}\frac{\partial r}{\partial y}, \quad H_y = +\frac{w}{c}\frac{e}{r^2}\frac{\partial r}{\partial x}, \quad H_z = 0 \quad \text{.........(1).}$$

We may therefore write the resultant magnetic intensity as

$$H = \frac{ew}{cr^2}\sin\theta \quad \text{........................(2).}$$

The magnetic force is evidently distributed in circles about the axis of motion.

The value of the magnetic force above is that for a charge e in uniform rectilinear motion. It will, however, give the instantaneous value of H in cases of curvilinear motion provided the acceleration is not too great (see Chap. XII). We shall suppose that this condition is satisfied by the intra-atomic motions which give rise to the magnetic quality of bodies. Let us seek an expression for the component in any direction of the magnetic intensity at any point Q due to an electron moving in a closed orbit.

Let it be required to find the component H_z of the magnetic intensity H at Q (Fig. 43). Resolve the velocity at every point of the orbit into two components, one parallel to Oz and the other perpendicular to Oz. The components parallel to Oz contribute nothing to the value of H_z. The other components can, for the whole orbit, be represented by the projection of the original orbit on a plane perpendicular to Oz. Let PMN be this projection. Let the velocity at P in the projected orbit be along PR, and take the origin O so that it lies in the plane of the original orbit. PMN and PR then lie in the plane xOy. Join PQ. The instantaneous resultant magnetic intensity H at Q due to a moving charge at P is perpendicular to the plane PQR. Let it be equal to QS and draw SZ perpendicular to Oz. Draw QR perpendicular to PR and join OR. Then all the lines OZ, ZS, SQ, QR and OR are perpendicular to PR. Moreover the angles

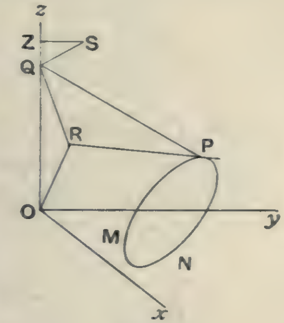


Fig. 43.

SZQ, SQR and QOR are right angles, so that the triangles ZQS, QRO are similar and the angles ZQS, QRO are equal. Thus

$$H_z/H = ZQ/QS = OR/RQ = p/r \sin\theta,$$

where p is the length of the perpendicular drawn from O on the projection of the line of motion, and r, θ are the coordinates of Q with respect to the point P as origin and the projection of the line of motion as axis. But we have seen that the resultant magnetic intensity H in the same notation is $\dfrac{ev \sin\theta}{cr^2}$, whence

$$H_z = \frac{epv}{cr^3} \text{ ......(3).}$$

We shall now apply this result to find the average value of the components of the magnetic force at any point P arising from the motion of an electron in an approximately circular orbit. Let QRS (Fig. 44) be the orbit, O its centre and ON the polar axis. Let the angle $PON = \theta$. We shall call the average component of magnetic intensity along OP the radial component and the average force at right angles to OP in the plane of the paper the tangential component.

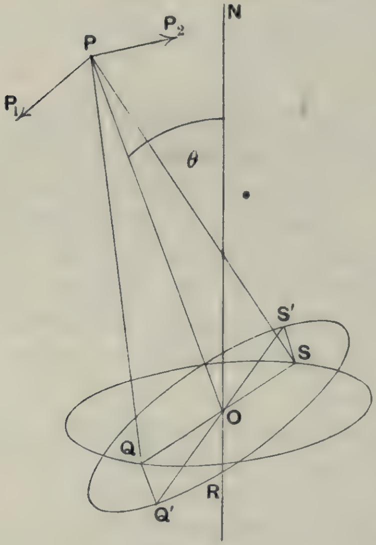


Fig. 44.

The Radial Component.

Consider the motion at any point Q of the orbit. It may be resolved into two parts, one parallel to OP and the other in the direction of the tangent at Q' to the projection $Q'RS'$ of QRS on a plane perpendicular to OP. The component of velocity parallel to OP contributes nothing to the radial component of the force at P. The direction of the force arising from the component of velocity in a direction perpendicular to OP will always be perpendicular to the radii QP, SP and so on. It will thus lie along the lines PP_1, PP_2 and so on. The radial component arising in this way will thus have the same sign as we proceed round the orbit. Provided the dimensions of the orbit are small enough it will evidently not matter if we take the moving point S to be at S', Q at Q' and so on. To the same order of accuracy we can treat the lines $PQ = PS = r$ as constant. Hence from (3) the average value of the force along OP is $H_r = \frac{e}{cr^3}\overline{pv}$. Now the average value $\overline{pv}$ is clearly equal to twice the area of the curve $Q'RS'$ divided by the periodic time τ of the orbit, p being the perpendicular from O on the tangent to this curve. Hence if S is the area of the original orbit QRS, $\overline{pv} = 2S\cos\theta/\tau$ and

$$H_r = \frac{2e}{cr^3}\,\frac{S}{\tau}\cos\theta \dots\dots\dots\dots\dots\dots\dots\dots (4).$$

The Tangential Component.

This is not so readily found, but a similar method of treatment may be applied with success. Take OP as the axis of z. The tangential component of force sought is the force in a direction PT (Fig. 45) perpendicular to OP in the plane PON. Call this the axis of y. The axis of x is thus perpendicular to the plane PON. Resolve the velocity at any point S of the orbit into its components $\dot{x}$, $\dot{y}$, $\dot{z}$. $\dot{y}$ being parallel to PT will contribute nothing to the tangential component at P. Let us consider the effects of the $\dot{z}$ and $\dot{x}$ components separately. Let the dotted curve represent the projection of the original orbit on a plane containing OP and perpendicular to the plane NOP. The dotted curve will thus be a representation of the $\dot{x}$ and $\dot{z}$ velocities and its area $= S\sin\theta$. Consider the $\dot{z}$ component of velocity first. The y component of magnetic force at P arising from this is always in the same direction whether the particle is above or below the plane of the paper, on account of the change of sign of the velocity.

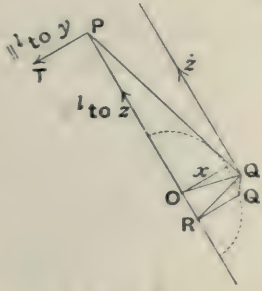


Fig. 45.

The magnetic force at P arising from this component of the motion at Q is equal to

$$\frac{e\dot{z}}{cQP^2}\sin QPR = \frac{e\dot{z}\,.\,QR}{cQP^3}.$$

This force is perpendicular to the plane POQ. What we want is the component along PT which is perpendicular to the plane POQ'. This is equal to

$$\frac{e\dot{z}QR}{c\,.\,QP^3}\times\frac{Q'R}{QR}=\frac{e\dot{z}Q'R}{c\,.\,QP^3},$$

since PR is the line of intersection of the planes QOP and $Q'OP$. On account of the effect being in the same sense all the way round the orbit we may to the first order in OQ/OP put

$$PQ = OP = r = \text{const.}$$

The component arising in this way is thus the average value of

$$\frac{e}{cr^3}\times Q'R\frac{dz}{dt}$$

or
$$\frac{1}{\tau}\frac{e}{cr^3}\int x\frac{dz}{dt}\,dt=\frac{e}{cr^3}\frac{S}{\tau}\sin\theta.$$

Now consider the y component of the force at P which arises from the x component of velocity in the orbit. Let VQR (Fig. 46) be the orbit, $V'Q'R'$ its projection in the plane of xz. Let SOU be the diameter perpendicular to OP and ON; this diameter will be perpendicular both to the element of the actual orbit at S and to its projection. Consider two points Q', R' symmetrically situated with respect to the axis OU. The corresponding points in the orbit are Q, R. It will be seen that the tangential force at P arising from the $\dot{x}$ motion at Q and R is oppositely directed at the two points. If the coordinates of R are x, y, z, those of Q are $x, -y, -z$. Moreover

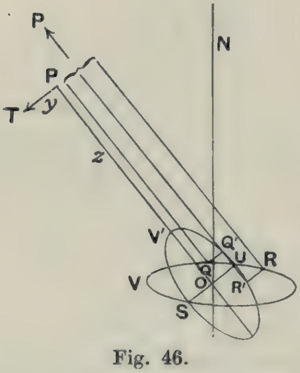


Fig. 46.

$$PR' = OP + z,\quad PQ' = OP - z.$$

If OP is sufficiently great compared with the dimensions of the orbit the average force in the y direction at P arising from the $\dot{x}$ motions at Q and R is

$$\frac{e\dot{x}}{2c}\left\{\frac{1}{(OP-z)^2}-\frac{1}{(OP+z)^2}\right\}=\frac{e\dot{x}}{2cOP^2}\left\{\left(1-\frac{z}{OP}\right)^{-2}-\left(1+\frac{z}{OP}\right)^{-2}\right\}$$
$$=\frac{e\dot{x}}{c\,.\,OP^2}\times\frac{2z}{OP}.$$

The average value of this taken all round the orbit is

$$2\,\frac{e}{c\,.\,OP^3}\,\frac{1}{\tau}\int z\,\frac{dx}{dt}\,dt=2\,\frac{e}{c\,.\,r^3}\,\frac{\text{area}\,SV'Q'R'}{\tau}=2\,\frac{e}{c\,.\,r^3}\,\frac{S}{\tau}\sin\theta.$$

The sign of this force is determined by the direction of the motion at the side of the orbit nearest to P and is evidently opposite to that arising from the component $\dot{z}$. The balance of tangential force is thus

$$H_t=\frac{e}{c\,.\,r^3}\,\frac{S}{\tau}\sin\theta \quad \text{....................}(5).$$

The average value of the component of the magnetic force perpendicular to both OP and PT vanishes. Because if AB is the line of intersection of the plane NOP with the plane of the orbit the latter can be divided into pairs of elements dS_1, dS_2 which are symmetrical about AB and are equidistant from P. These elements produce equal and opposite effects at P so far as the component under consideration is concerned. The orbit can thus be divided into mutually interfering pairs of points so that the average value of this component is zero.

It is evident from formulae (4) and (5) that the average value of the magnetic field of force due to the revolving electron is exactly equivalent to that which would be given by a small magnet whose moment is

$$M=\frac{eS}{c\tau} \quad \text{..........................}(6)$$

and whose axis coincides with the axis of revolution of the electron.

An atom may in general contain a number of electrons rotating in closed orbits as well as others which execute small oscillations about a position of static equilibrium. The orbits may be numerous and distributed in various azimuths inside the atom, so as to furnish no resultant magnetic moment; or they may possess an axis of symmetry with a resultant magnetic moment. In every case we shall see that a phenomenon analogous to diamagnetism occurs; but in the case of the atoms for which the revolving electrons have a resultant magnetic moment there are reasons for believing that this is marked by the paramagnetic or ferromagnetic effects which supervene.

In developing a theory of the action of an external magnetic field on the revolving electrons it is necessary to make some hypothesis about the nature of the forces which hold them in their orbits and which determine the orientation of the orbits with respect to the atoms. We shall suppose that the forces are determined by the structure of the atom and that the planes of the orbits are determined by the symmetry of the atom. When an external field is applied, forces are brought to bear on the revolving electrons which derange the previous state of motion. This displacement will give rise to a force acting on the neighbouring parts of the atom which will, in general, cause the axis of the atom to turn and so change the plane of the orbit. The state of things we are imagining is in fact much the same as if the electron were revolving in a channel cut in a rigid non-conductor (the atom). If in the absence of an external magnetic field the orbits are arranged so that the atom has no magnetic axis, these forces will not give rise to any tendency to change the orientation of the atom as a whole. In such cases we shall see that the effects produced give rise to phenomena like those exhibited by diamagnetic substances. The same results would follow if the atoms were held rigid by interatomic constraints; but as a number of liquids are diamagnetic such a supposition would not help in explaining diamagnetism. The hypothesis that the revolving electrons can be treated like currents flowing round their orbits, which are more or less rigidly attached to the atom, simplifies the mathematical calculations very considerably.

Diamagnetism.

Let us suppose that, for reasons of symmetry or otherwise, the external field has no tendency to alter the orientation of the orbits in space. We can calculate the magnetic permeability of a substance if we can calculate the change in the equivalent moment of the revolving electrons produced by the application of a given external field H. Let S' be the area of the orbit projected on a plane perpendicular to H. Then the establishment of an external field H will cause a flux HS' of magnetic force through the orbit. If the orbit is circular this will give rise to an electric intensity E tangential to the orbit, where

$$\int E\,ds = -\frac{1}{c}\frac{d}{dt}(HS\cos\theta) = -\frac{\delta(HS\cos\theta)}{c\tau}$$

$$= -\frac{\omega}{2\pi c}\,\delta(HS\cos\theta) \quad\text{..........................(7)},$$

where θ is the angle between H and the normal to S, τ is the periodic time, ω the angular velocity and δ denotes the change per revolution.

But the moment of the force Ee will cause an increase in the moment of momentum of the electron. Thus

$$Eer = \frac{d}{dt}(mr^2\omega) = 2\,\frac{mc}{e}\,\frac{dM}{dt},$$

since $$M = \frac{eS}{c\tau} = \frac{e}{c}\,\frac{\omega r^2}{2}.$$

So that $$dM = \frac{e^2}{2mc}\,Er\,dt = \frac{e^2}{2mc\omega}\,E\,ds,$$

and $$\delta M = \frac{e^2}{2mc\omega}\int E\,ds = -\frac{e^2}{4\pi mc^2}\,\delta(HS\cos\theta),$$

neglecting the change in ω during one revolution. Hence

$$\frac{\delta M}{M} = -\frac{\tau}{4\pi}\,\frac{e}{mc}\,\frac{\delta(HS\cos\theta)}{S} \quad\text{..................(8)}.$$

If we neglect the change in S compared with S, and if we take $\tau = 10^{-15}$ sec. to correspond with spectral lines, since

$$\frac{e}{m} = 1{\cdot}77 \times 10^7 \times c\sqrt{4\pi},$$

$\delta M/M$ is of the order $10^{-9} \times \delta H$. Since the greatest attainable magnetic fields correspond to $H < 10^5$ it follows that the effect we are considering will not change the magnetic moment of the orbits by more than about 10^{-4} of its value.

We shall now consider the conditions under which the field $H \cos \theta$ will change the area S of the orbit, limiting ourselves to the case of a circular orbit under a central force. If this is $f(r)$ at distance r, then the condition for steady motion in the absence of the magnetic field is

$$m\omega^2 r = f(r).$$

After the magnetic field is applied this becomes

$$m\omega^2 r + m\omega^2 \delta r + 2m\omega r\delta\omega + \frac{H}{c}\cos\theta\, e\omega r = f(r) + f'(r)\,\delta r,$$

neglecting squares and higher powers of δr and $\delta\omega$. Thus

$$\{f'(r) - m\omega^2\}\,\delta r = 2mr\omega\,\delta\omega + \frac{e}{c}\,\omega r\, H\cos\theta.$$

But

$$\delta\left(\tfrac{1}{2}\omega r^2\right) = \frac{c\delta M}{e} = -\frac{e}{4\pi mc}\,\delta\left(HS\cos\theta\right) = -\frac{er^2}{4mc}\,H\cos\theta,$$

since we are neglecting terms involving the product of H and δr as small. Thus

$$-4m\omega^2\delta r = 2m\omega r\delta\omega + \frac{e}{c}\,\omega r\, H\cos\theta.$$

Whence $$\{f'(r) + 3m\omega^2\}\,\delta r = 0.$$

It follows that either

$$(1)\quad \delta r = 0 \quad\text{and}\quad \delta\omega = -\frac{He\cos\theta}{2mc},$$

or $$(2)\quad f'(r) = -3m\omega^2 \quad\text{or}\quad f'(r)/f(r) = -3/r,$$

and $$f(r) = \text{constant} \times r^{-3}.$$

Thus, except in the special case in which the force varies inversely as the cube of the distance, $\delta r = 0$, and there will be no change in the area of the orbit. The only effect of the magnetic field will be to change the angular velocity.

It is clear that the component of the magnetic intensity $H \sin\theta$, in the plane of the orbit, will not change the area of the latter, as it only gives rise to displacements perpendicular to this plane.

Calculation of the Magnetic Permeability.

The effect that we have been considering involves a *diminution* in the component of the magnetic moment of the elementary orbits resolved along the direction of the external magnetic field. The phenomenon is, in fact, precisely analogous to that of the induction of currents in linear conductors occupying the same positions as the orbits. The net result is a creation of polarization of the elements of the medium in a direction opposite to that of H. Since the creation of positive polarization in the electrostatic case leads to a dielectric constant greater than unity it is clear that the present effect will lead to a magnetic permeability less than unity. In other words the effect we are now considering leads to diamagnetism and not to paramagnetism. The value of the diamagnetic constant can readily be calculated.

We have seen that the increment in M per orbit due to the establishment of H is

$$-\frac{e^2}{4\pi mc^2} H'S,$$

where $H' = H\cos\theta$ is the component of H along the normal to the orbit. Suppose that the atoms of the body considered possess no resultant magnetic moment, then the only effect of the field H will be to produce this change δM. Suppose that there are ν atoms per unit volume, each of which contains n electrons executing orbits whose areas are $S_1, S_2, \ldots S_n$. Let the normal ON_p to the plane of any orbit make an angle θ_p with the direction of H. On the average all directions are equally probable for the line ON_p, so that, out of any number of orbits considered, the proportion which lie between θ and $\theta + d\theta$ will be $\dfrac{2\pi\sin\theta\,d\theta}{4\pi}$. For each orbit of type p

$$\delta M_p = -\frac{e^2 S_p}{4\pi mc^2} H\cos\theta_p,$$

and to get the resolved part of this parallel to H we have to multiply again by $\cos\theta_p$. Thus if there are ν_p orbits of this type in unit volume their contribution to the magnetic moment of the medium will be

$$\begin{aligned} M_p &= -\frac{e^2 H}{8\pi mc^2}\nu_p S_p \int_0^\pi \sin\theta\cos^2\theta\,d\theta \\ &= -\frac{1}{3}\frac{e^2 H}{4\pi mc^2}\nu_p S_p. \end{aligned}$$

To get the intensity of the polarization I we have to sum this for all the different orbits in the atom. Thus

$$I = \sum_{p=1}^{p=n} M_p = -\frac{1}{3}\frac{e^2}{4\pi mc^2} H \sum_{p=1}^{n} \nu_p S_p.$$

The force in a cavity perpendicular to the lines of force is

$$B = \mu H = H + I = H\left(1 - \frac{e^2 \sum_1^n \nu_p S_p}{12\pi mc^2}\right),$$

so that
$$\mu = 1 - \frac{e^2 \sum_1^n \nu_p S_p}{12\pi mc^2},$$

and
$$1 - \mu = \frac{e^2 \sum_1^n \nu_p S_p}{12\pi mc^2} \quad \text{.......................(9).}$$

It is necessary to show that the known values of $1-\mu$ do not lead to absurd values of S_p the areas of the orbits. Of the known substances bismuth has the largest value of $1-\mu$, namely $3{\cdot}1 \times 10^{-5}$. Let us suppose each atom to contain one orbit of each kind, then ν_p becomes ν the number of atoms in unit volume. We can now estimate ΣS_p the sum of the areas of all the orbits in the atom. As a sufficient approximation we shall take

$$e/m = 1{\cdot}8 \times 10^7 \times c\sqrt{4\pi}, \quad e = 10^{-20} \times c\sqrt{4\pi},$$

and
$$\nu = \frac{9{\cdot}78}{207} \times 10^{24} = 5 \times 10^{22},$$

9·78 being the density of bismuth, 207 its atomic weight, and 10^{-24} the mass of the hydrogen atom. Thus $\sum_1^n S_p = 10^{-14}$ for bismuth. If the area of the orbits were comparable with the cross section of the atom we should have $S = \pi \times 10^{-16}$, so that, treating all the orbits as of equal area, $n = 30$ approximately. In the case of other substances, whose diamagnetism is less pronounced, the values of n would be smaller than this. The atomic weight of bismuth is 207, and there are good reasons for believing that the number of electrons in each atom of the different elements is comparable with the atomic weight. It is thus clear that the diamagnetic coefficients are of the order of magnitude which is otherwise probable.

We have seen that the effect of applying a magnetic field is to change the periods of the orbits projected into a plane perpendicular to the magnetic field, the component of the motion parallel to the latter being unaltered. Since the revolving electrons are accelerated they will in general be radiating, and the frequencies of the radiation will be determined by the periods of revolution. There will thus be a change in the frequency of the radiation produced by the magnetic field. The displacement of the spectral lines due to this cause was discovered by Zeeman in 1896. The foregoing method is not a satisfactory one for determining quantitatively the changes in the frequency and character of the radiation which arise in this way, as it does not sufficiently consider the motions perpendicular to the orbit. This deficiency will be remedied when the theory of the Zeeman effect is considered in Chapter XX. Without going more deeply into the matter at this stage, it is evident from what has been said, that the magnetic displacement of the spectral lines and the phenomenon of diamagnetism are very intimately related, on the theory we are discussing.

It appears that the occurrence of electrons revolving in orbits is quite unnecessary to account for diamagnetism. The same kind of effects occur even if the electrons are at rest before the magnetic field is applied. This is shown very clearly by the following investigation which is due to Lorentz*.

We shall confine ourselves to arrangements of electrons which are isotropic with respect to three mutually perpendicular directions. Let the coordinates of any particular electron with respect to any set of rectangular axes, whose origin O is the centre of mass of the system, be x, y, z. Then $\Sigma x = \Sigma y = \Sigma z = 0$, taken over the whole body. Let its moment of inertia about any axis through O be $I = 2mk$. Then

$$k = \Sigma x^2 = \Sigma y^2 = \Sigma z^2 \quad \text{and} \quad \Sigma xy = \Sigma xz = \Sigma yz = 0.$$

Let E be the resultant electromotive intensity of external origin at any point, then the force acting on an electron has the components eE_x, eE_y, eE_z, and the couple about O has the components

$$e\Sigma(yE_z - zE_y), \quad e\Sigma(zE_x - xE_z), \quad e\Sigma(xE_y - yE_x).$$

---

* *Theory of Electrons*, p. 124.

If the whole system is very small E will not vary much from point to point of it, so that we can put as a sufficient approximation:

$$E_x = F_x + x\frac{\partial F_x}{\partial x} + y\frac{\partial F_x}{\partial y} + z\frac{\partial F_x}{\partial z},$$

$$E_y = F_y + x\frac{\partial F_y}{\partial x} + y\frac{\partial F_y}{\partial y} + z\frac{\partial F_y}{\partial z},$$

$$E_z = F_z + x\frac{\partial F_z}{\partial x} + y\frac{\partial F_z}{\partial y} + z\frac{\partial F_z}{\partial z},$$

where F is the value of E at the origin. Hence since Σx and Σxy, etc. vanish, the components of the couple become

$$ek\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right),\quad ek\left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right),\quad ek\left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right),$$

or

$$-\frac{e}{c}k\dot{H}_x,\quad -\frac{e}{c}k\dot{H}_y,\quad -\frac{e}{c}k\dot{H}_z.$$

There is thus an angular acceleration about the axis of $H = -\frac{e}{2mc}\dot{H}$, and the creation of a field H therefore results in the creation of angular velocity $-\frac{e}{2mc}H$.

If we take the system considered by Lorentz to be one of the atoms of the substance we see that the effect of placing it in a magnetic field of strength H will be to set all the electrons in the atom in rotation about the axis with the uniform angular velocity $-\frac{e}{2mc}H$. This rotation, whose axis is parallel to H, will give rise to an intensity of magnetization in the same sense as that given by Langevin's theory. Moreover the magnitudes are the same in both cases provided we replace $S\cos\theta$ by πr^2, where r is the distance of an electron from an axis, parallel to H, which passes through the centre of figure of the atom. The resulting value of the permeability is easily shown to be

$$\mu = 1 - \frac{e^2\nu\,\Sigma\, r^2}{8mc^2}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10),$$

where ν is the number of atoms in unit volume, and the summation is extended over all the electrons in the atom. The order of magnitude of $1-\mu$ is evidently the same as before, and is equally in agreement with estimates of the number of electrons in the atom which are derived from other sources*.

The theory outlined brings diamagnetism into agreement with the fact that the Zeeman separation results in sharp lines. The alteration in the frequency of the emitted radiation depends only on the magnitude of H and is independent of its direction. In general the Zeeman effect is much more complicated than that outlined by this theory which, nevertheless, probably contains the gist of the matter.

Two important deductions about diamagnetism may be drawn if it is admitted to be identical in nature with the phenomenon recognized optically as the Zeeman effect. In the first place the Zeeman effect is exhibited by practically all the spectral lines of every substance, so that we should expect every substance to have diamagnetic quality. This does not really involve any contradiction with experience, as the diamagnetic property is necessarily very feeble, and is therefore easily masked by small paramagnetic effects. There seems to be no valid reason for supposing that the same kind of actions which produce the diamagnetism of bismuth do not occur and produce similar effects even in substances like iron.

The other point is that $\mu - 1$ is proportional to a universal constant $-\frac{e^2}{8mc^2}$ multiplied by $\nu \Sigma r^2$. Now, provided the unit of symmetry we have considered is the atom, we should expect ν and Σr^2 both to be independent of temperature within the order of accuracy to which $\mu - 1$ can be measured. Of the substances for which the variation of $\mu - 1$ with the temperature was measured by Curie†, water, quartz, KNO_3 and molten bismuth showed no detectable variation. In the case of solid bismuth on the other hand the value of $1 - \mu$, which was large at ordinary temperatures, fell off in a linear manner as the temperature increased to the melting point, when there was a sudden drop to the small value characteristic of the fused metal. It seems probable that these changes are connected with the crystalline structure of the substance.

---

* It is not necessary that all the electrons in the unit considered should be able to rotate about the axis of symmetry. The momentum calculated above will be communicated to all of them, but some may be prevented from rotating by constitutive restoring forces. The value of the diamagnetic constant may therefore only indicate a lower limit for the number of electrons in the atom.

† *Oeuvres*, p. 252.

Paramagnetism.

To explain the magnetic qualities of bodies, other than diamagnetic, we shall make the hypothesis that, in some cases at any rate, when the resultant magnetic moment of the atom is not zero, the applied magnetic field is able to turn the planes of the orbits of the constituent electrons of the atoms towards the plane perpendicular to its direction. We do not know the precise nature of the mechanism by which the turning is brought about; but in defence of the hypothesis we are able to urge that such a rotation tends to make the potential energy of the system a minimum, and will therefore tend to occur if there is any means by which it can be accomplished. Superposed on this there will in every case be the diamagnetic effects already discussed; in many cases, however, these are insignificant compared with the effects which arise from the turning of the orbits.

The couple which tends to turn the axis of an orbit depends on the mutual energy of the external magnetic field H_e and that of the revolving electron. Let the magnetic force H_i, due to the revolving electron at any point, consist of two parts, an average value H_1 and an oscillating part H_2. Then $\overline{H_i} = H_1$ and $\overline{H_2} = 0$, where the bars denote mean values of the vectors taken over a complete revolution. The energy per unit volume at any point is

$$\tfrac{1}{2}(H_e + H_1 + H_2)^2 = \tfrac{1}{2}\{H_e^2 + H_1^2 + H_2^2 + 2(H_e H_1) + 2(H_e H_2) + 2(H_1 H_2)\},$$

where $(H_e + H_1 + H_2)$ denotes the vector sum of these vectors.

The mean value of this taken over a revolution is

$$\tfrac{1}{2}\{\overline{H_e^2} + \overline{H_1^2} + \overline{H_2^2} + 2(H_e H_1)\},$$

since the average of the other terms is zero. Thus the mean value of the mutual energy, per unit volume of the medium, is $(H_e H_1)$, and does not depend on H_2. It follows that the forces which on the average produce a given rotation of the planes of the orbits will be the same as those which would produce the same effect on a similarly situated equivalent current. The revolving electrons can therefore be replaced by the equivalent currents so far as the rotational effect of an external magnetic field is concerned.

If there were no other actions than those arising from the magnetic fields, the orbits would all set themselves with their normals parallel to the external magnetic force, as this is the arrangement in which the potential energy is a minimum. The tendency to complete alignment will however be resisted by the thermal motions of the atoms and also by the interatomic forces of other than magnetic type. The latter embrace the forces which give rise to cohesion and to chemical effects and which in all probability are mainly electrostatic. One effect of the application of a magnetic field will be to convert the mutual potential energy of the orbit and the field into kinetic energy of the matter which moves with the orbit. In the simplest case, as perhaps in a gas where the elastic and chemical forces may be neglected, the mutual energy will be entirely transformed into the kinetic form. This, however, will change the distribution of kinetic energy among the different atoms or sub-atoms so that it is no longer that which was characteristic of the substance at the original temperature. One of the effects of magnetization then may be that of changing the temperature of the substance.

If the magnetizable substance is in a field of strength H and its intensity of magnetization is changed from I to $I + dI$, the work done by the magnetic force in increasing the magnetization is HdI. If heat dQ is communicated to the substance it will partly be used in changing that part of the internal energy U of the substance which does not depend directly on the intensity of magnetization, and partly in altering the latter. Thus we shall have

$$dQ = dU + HdI.$$

In general U and I may be functions of H as well as of the absolute temperature T. For processes which are reversible, that is to say, where there is no hysteresis, the increment of entropy

$$dS = \frac{dQ}{T} = \frac{1}{T}\left(\frac{\partial U}{\partial H} + H\frac{\partial I}{\partial H}\right) dH + \frac{1}{T}\left(\frac{\partial U}{\partial T} + H\frac{\partial I}{\partial T}\right) dT$$

must be a perfect differential. Thus

$$\frac{\partial}{\partial H}\left\{\frac{1}{T}\left(\frac{\partial U}{\partial T}+H\frac{\partial I}{\partial T}\right)\right\}=\frac{\partial}{\partial T}\left\{\frac{1}{T}\left(\frac{\partial U}{\partial H}+H\frac{\partial I}{\partial H}\right)\right\}$$

or

$$\frac{\partial I}{\partial T}+\frac{H}{T}\frac{\partial I}{\partial H}+\frac{1}{T}\frac{\partial U}{\partial H}=0.$$

In general U will probably involve H on account of the strains produced in the material by the magnetic field. With gases $\frac{1}{T}\frac{\partial U}{\partial H}$ will be zero and in most cases it will probably be very small; so that we obtain as an approximation covering a majority of cases, the equation

$$\frac{\partial I}{\partial T}+\frac{H}{T}\frac{\partial I}{\partial H}=0 \quad \text{.....................(11).}$$

This is satisfied by $I=f(H/T)$, where f is any function.

In the case of paramagnetic substances I is directly proportional to H, so that, if the conditions above are satisfied,

$$I=\frac{A}{T}H \quad \text{...........................(12),}$$

where A is a constant independent of T. Curie found that for a number of typical paramagnetic substances the susceptibility I/H was inversely proportional to the absolute temperature. For such substances it would seem to follow that U does not depend appreciably on H, that the only important change produced by H is in the potential energy of the elementary magnets and that the energy of the accompanying strains, if any, is negligible.

The result contained in equation (12) was first discovered by experiment, and is often referred to as Curie's Law. The quantity A is also sometimes called Curie's constant. A considerable number of exceptions to the law have been found*, particularly at low temperatures. Some of these have been attributed to changes in the crystalline or other configuration of the material.

---

* Cf. Kammerlingh Onnes and Perrier, *Konink. Akad. Wetensch. Amsterdam Proc.* vol. xiv. p. 115 (1911).

Paramagnetic Gas.

An instructive case, which has been considered by Langevin, is that of a paramagnetic gas such as oxygen. In this case the kinetic theory of gases enables us to calculate the form of the function f in the relation $I = f\left(\frac{H}{T}\right)$ which is required by the laws of thermodynamics. We know from Boltzmann's Theorem (see Chap. XVII, p. 403) that if the molecules of a gas in a closed vessel occupy positions in which they have varying amounts of potential energy, then there will be a greater number per unit volume in the positions in which the potential energy is less. In fact the ratio of the concentrations of the gas at two points where the potential energy differs by w is $e^{-w/RT}$, where R is the gas constant for a single molecule.

In the present instance w will be the potential energy of the equivalent magnet in the field H, i.e. $-MH\cos\theta$, where θ is the angle between the magnetic axis of the molecule whose moment is M and the field H. The number of molecules dn whose magnetic axes lie within the two cones whose semi-angles are θ and $\theta + d\theta$ respectively is

$$dn = 2\pi A e^{\frac{MH}{RT}\cos\theta} \sin\theta \, d\theta,$$

where A is a constant as yet undetermined. The total number N of molecules considered is evidently

$$\begin{aligned} N &= 2\pi A \int_0^\pi e^{\frac{MH}{RT}\cos\theta} \sin\theta \, d\theta \\ &= 2\pi A \frac{RT}{MH}\left\{e^{\frac{MH}{RT}} - e^{-\frac{MH}{RT}}\right\}. \end{aligned}$$

Thus $$A = \frac{N}{4\pi}\frac{MH}{RT}\Big/\sinh\frac{MH}{RT}.$$

The resultant intensity of magnetization I is in the direction of H, by symmetry, and is given by

$$\begin{aligned} I &= \int M\cos\theta \, dn = 2\pi A M \int_{-1}^{+1} x e^{\frac{MH}{RT}x} \, dx \\ &= 4\pi A M\left\{\frac{\cosh a}{a} - \frac{\sinh a}{a^2}\right\}, \end{aligned}$$

where $a = MH/RT$. But $A = Na/4\pi \sinh a$, so that

$$I = MN\left(\frac{\cosh a}{\sinh a} - \frac{1}{a}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots(13).$$

Since M is determined by the structure of the molecules, we see that for a given density of gas I depends only on a, i.e. I is a function only of H/T in accordance with the conclusion already reached by thermodynamic reasoning. We also note that I is proportional to N, i.e. to the pressure of the gas.

It has not been possible as yet to test this formula by experiments on gases on account of the smallness of the intensities of magnetization which they acquire. It has, however, been extensively used by P. Weiss in building up a theory of the behaviour of ferromagnetic substances, as we shall see.

Ferromagnetic Substances.

The main difference between the ferromagnetic and the paramagnetic substances lies in the very high intensities of magnetization attainable by the former, in the fact that they are capable of permanent magnetization, and that the magnetization is not in general a definite function of the external magnetic force H. The magnetization does not change reversibly with H and substances of this class therefore exhibit the phenomenon of hysteresis. It results that the heat $Q = \int H\,dI$, which is developed when H is made to pass through a cycle of changes, no longer vanishes but has a finite value.

Weiss* has attempted to explain the facts of ferromagnetism on the hypothesis that the only forces which it is necessary to consider as acting on the elementary magnets or revolving electrons are (1) the impressed magnetic field H_1 of external origin, and (2) the molecular field H_2 arising from the elementary magnets of the neighbouring atoms. It is also necessary to take into account the kinetic reactions arising from the thermal agitation of the molecules, just as in Langevin's theory of a paramagnetic gas.

The assumption of a molecular field uniform throughout the substance will naturally be no more accurate than the similar assumptions as to uniformity which are almost invariably made in dealing with molecular physics. This field will vary greatly from point to point but we may reasonably expect to get valuable indications of the way in which a real body would behave by assuming for H_2 a uniform value equal to its average value throughout the substance. The theory entirely omits to consider interatomic forces of non-magnetic type. This is clearly legitimate until it is shown that such forces do play an important part in the phenomena under consideration.

---

* *Journal de Physique*, vol. VI. p. 661 (1907).

Permanent Magnetization.

Weiss's theory accounts for the existence of permanent magnetization in the following fashion. Bearing in mind that the molecular magnets are in equilibrium under the influence of thermal agitation and the intensities H_1 and H_2, the value of I for a given value of H will be determined by the equations

$$\frac{I}{I_0} = \frac{\cosh a}{\sinh a} - \frac{1}{a}, \quad a = \frac{MH}{RT} \quad \text{and} \quad H = H_1 + H_2,$$

where $I_0 = MN$ is the maximum possible intensity of magnetization which is attained when the axes of the elementary magnets all point in the same direction. Moreover H_2 is proportional to the intensity of magnetization and may be written $H_2 = \lambda I$, where λ is some constant. Permanent magnetization corresponds to the absence of external field, so that $H_1 = 0$, and we have the two independent equations

$$\frac{I}{I_0} = \frac{\cosh a}{\sinh a} - \frac{1}{a} \quad \text{and} \quad \frac{I}{I_0} = \frac{RT}{\lambda M I_0} a \text{............} (14).$$

The values of I which satisfy these equations may be found most readily by a graphical method (see Fig. 47). Let ORP be the graph of $\frac{I}{I_0} = \frac{\cosh a}{\sinh a} - \frac{1}{a}$ and OQP that of $\frac{I}{I_0} = \frac{RT}{\lambda M I_0} a$. According to the simplest hypothesis which can be made R, λ, M and I_0 are constants independent of T and H, so that OQP is a straight line which makes an angle α with the axis of x, where $\tan \alpha = \frac{RT}{M\lambda I_0}$ is proportional to the absolute temperature T. The possible values of I/I_0 are given by the intersections of OQP and ORP. There are thus always two possible values of I, one of which is zero. Of these, however, the point P corresponds to a stable and the point O to an unstable condition of the substance. To show this, let us suppose that by means of an external field the intensity of magnetization is made to undergo a slight decrease, so that it is now determined by the line ST parallel to the axis of a. This state of magnetization gives rise to an internal field which is proportional to OU, whereas to overcome the thermal tendency to disorganization it is only necessary to have a field proportional to OV. The magnetization of the substance will therefore increase automatically until the state corresponding to the point P is reached. The reverse happens if the magnetization is given a virtual increase beyond that which corresponds to the point P. Thus P represents a stable configuration of the material. In the same way it can be shown that O represents an unstable condition.

The permanent magnetization exhibited by the paramagnetic metals is not as definite as this theory would lead one to expect.

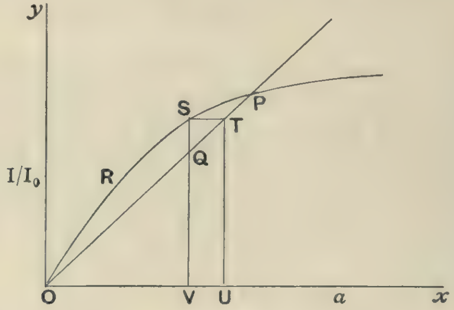


Fig. 47.

The indefiniteness may, however, be due to the fact that these materials are not microscopically homogeneous, as well as to the occurrence of microscopic local reversals of magnetization. There are also, in all probability, complications arising from the crystalline character of these materials.

As the temperature rises the slope of the line OP increases until at a certain temperature T_0 it coincides with the tangent to the curve at the origin. At higher temperatures than this the only possible solution would be $I=0$, so that the substance would be incapable of permanent magnetization in the absence of an external magnetic field. The temperature T_0 may therefore be interpreted as that at which the ferromagnetic quality disappears.

To determine the slope of the tangent to the curve at the origin we have

$$\frac{d}{da}\left(\frac{I}{I_0}\right)=\frac{d}{da}\left(\frac{\cosh a}{\sinh a}-\frac{1}{a}\right)$$

$$=1-\left(\frac{e^a+e^{-a}}{e^a-e^{-a}}\right)^2+\frac{1}{a^2}$$

$$=1-\frac{1}{a^2}\left(1+\frac{a^2}{2}\right)^2\left(1+\frac{a^2}{6}\right)^{-2}+\frac{1}{a^2}$$

$$=\tfrac{1}{3} \text{ when } a=0.$$

Thus
$$\frac{1}{3}=\frac{d}{da}\left(\frac{I}{I_0}\right)_0=\left(\frac{I}{I_0 a}\right)_{a=0}=\frac{RT_0}{\lambda M I_0},$$

and
$$T_0=\frac{\lambda M I_0}{3R}\text{ .... }(15).$$

On this theory if we write $f(a)$ for $\frac{\cosh a}{\sinh a}-\frac{1}{a}$ we have

$$I/I_0=f(MH_2/RT),$$

where f is the same for all substances. Since $H_2=\lambda I$ we have

$$\frac{I}{I_0}=f\left(\frac{\lambda MI}{RT}\right)=f\left(3\frac{T_0}{T}\frac{I}{I_0}\right),$$

so that
$$I/I_0=\phi(T/T_0)\text{ .... }(16),$$

where the function ϕ is the same for all substances. Thus if we express I, the intensity of permanent magnetization, in terms of the maximum possible intensity of magnetization I_0, and T the absolute temperature in terms of the absolute critical temperature T_0, we obtain a characteristic equation for the intensity of permanent magnetization which is identical for all ferromagnetic substances.

It is probable, as we shall see later, that the magnetic properties of the ferromagnetic metals are too much complicated by various secondary causes to afford a satisfactory test of the above theoretical conclusions. The properties of various crystalline ferromagnetic minerals, though complex enough of themselves, are in some respects simpler than those of the ferromagnetic metals and are better suited for carrying out a test of this character. The mineral magnetite has been found to be especially suitable for this purpose; but before discussing the experimental data which have been obtained with this substance we shall consider briefly the remarkable magnetic properties of another mineral, pyrrhotite, which seem likely to shed much light on the nature of ferromagnetic substances in general.

The Properties of Pyrrhotite.

The magnetic properties of this ferromagnetic mineral, which is a sulphide of iron having a composition very near to FeS, but with a slight excess of sulphur, have been investigated in detail by Weiss. The crystal has three mutually perpendicular planes of symmetry which may be indicated by the axes Ox, Oy, Oz. The crystals are much more easily magnetized parallel to one of these axes Ox than in any other direction, and furthermore the susceptibility parallel to Oy is much greater than that parallel to Oz. The plane xOy is called by Weiss the magnetic plane.

The magnetic phenomena exhibited by a uniform crystal when placed in a magnetic field parallel to Ox are characterized by remarkable simplicity. If the crystal shows no magnetic polarity to start with, the intensity of magnetization remains zero until H reaches a critical value $+H_c$, when the intensity of magnetization suddenly assumes its saturation value $+I_s$, which remains constant for all positive values of H and for all negative values greater than $-H_c$. As soon as $-H_c$ is reached the intensity of magnetization suddenly becomes $-I_s$, and retains that value until the field becomes $=$ or $> +H_c$. The I, H curve is thus a rectangle and the phenomenon is irreversible.

If the magnetizing field is inclined to Ox the phenomena are very different*. For different values of H the curves for I obtained by rotating the direction of H in the magnetic plane are shown in the accompanying figure. If H exceeds about 12,000 gauss the value of I is constant for all values of the field, but unless H is very great I is not in the same direction as H. This effect is well exhibited by the curve for $H = 4000$. The short lines represent the *direction* of the resultant magnetization for a field of intensity $H = 4000$ gauss inclined to the axis Ox at an angle given in degrees by the numbers alongside. For a variation in H of 5°

---

* P. Weiss, *Journal de Physique*, vol. IV. p. 486 (1905).

from the axis *Oy* the direction of the intensity of magnetization varies more than 45°. The component of the intensity of magnetization parallel to the field is a minimum when the angle θ between H and Ox is a maximum, for small fields. For larger fields the minimum occurs at intermediate values*.

The curves in Fig. 48 can be represented quite closely by a simple trigonometrical formula. If θ is the angle between the axis Ox and H, and ϕ that between Ox and I, then it appears that

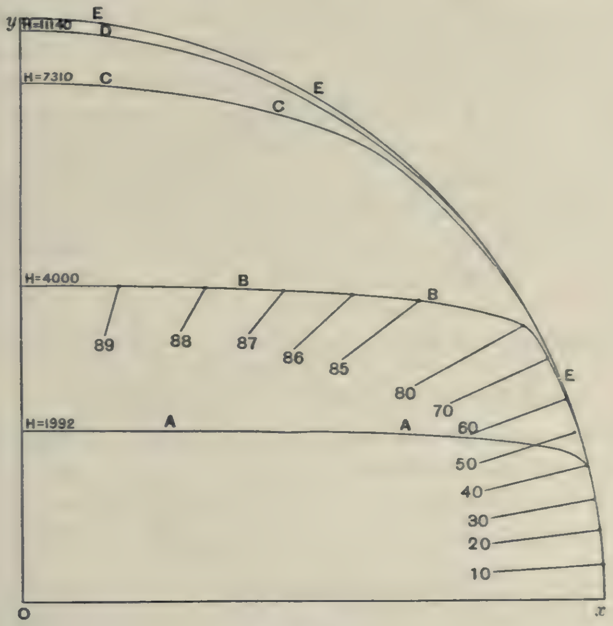


Fig. 48.

$$H \sin(\theta - \phi) - nI \sin\phi \cos\phi = 0 \;\ldots\ldots\ldots\ldots(17),$$

where n is a constant quantity. The phenomena in the plane xOz may be represented by a similar formula but with a different value of n.

A simple physical interpretation can be given to the foregoing equation. In addition to the field H the elementary magnets will be acted on by forces which depend upon the magnetization of the

---

[*] P. Weiss, *loc. cit.* p. 487.

medium. These forces will certainly be partly magnetic, arising from the intermolecular magnetic field. They will not, at any rate necessarily, be entirely magnetic. They may arise as reactions, of elastic or other type, to the displacement of the elements of the material, from the normal equilibrium position, which is produced by the magnetic field. Since I measures a displacement of the elements from either a more normal or a less regular arrangement, the potential energy of the system which is due to these forces will be proportional to I^2 even though the forces are not actually magnetic. On account of the aeolotropy of the medium the coefficient of I^2 will be different along the different axes. Thus in general when H and I make angles θ and ϕ respectively with the axis of x, the x component of force per unit magnetic moment of an elementary magnet may be represented by $H \cos \theta + \lambda_1 I \cos \phi$ and the y component by $H \sin \theta + \lambda_2 I \sin \phi$, where λ_1 and λ_2 are constants depending on the structure of the material. Thus taking moments about the centre of mass of the elementary magnet

$$(H \cos \theta + \lambda_1 I \cos \phi) \sin \phi = (H \sin \theta + \lambda_2 I \sin \phi) \cos \phi,$$

or

$$H \sin (\theta - \phi) - (\lambda_1 - \lambda_2) I \sin \phi \cos \phi = 0.$$

This is the same as the former equation if $\lambda_1 - \lambda_2 = n$. For the plane xOz we have only to replace $\lambda_1 - \lambda_2$ by $\lambda_1 - \lambda_3$. It appears from the experiments that $(\lambda_1 - \lambda_2) I = 7300$ gauss and

$$(\lambda_1 - \lambda_3) I = 150{,}000 \text{ gauss},$$

whereas the maximum (saturation) value of I is about 47 gauss. The great difference between these numbers is somewhat surprising. If they were really magnetic forces one might expect them all to be of the same order of magnitude.

One of the most striking features of these phenomena is the very small field which can reverse the magnetization along the axis Ox compared with the fields which are required to produce any appreciable magnetization along the perpendicular axes. This is the more striking as the reversal of the magnetization would appear to involve the intermediate passage of the elementary magnets through the perpendicular orientation which is so difficult to produce, throughout the mass of the material, at any rate, by the application of an external field. There is not, however, any essential difficulty here. It seems clear that there are in general two stable positions of the elementary magnets, namely the positive and negative directions along the axis of x. In the presence of an external field which exceeds $\pm H_c$ in magnitude, one of these becomes unstable. As the equilibrium is a kinetic and statistical one, as is shown (see p. 388) by the variation of the intensity of permanent magnetization with the temperature, there will be a continuous passage between the two states, so that in a very short time the molecules will have arranged themselves with the axes of the magnetic atoms all oscillating about the position of stable equilibrium. This condition will ultimately be reached, no matter how great the forces which oppose the intervening motion.

Permanent Magnetization and Temperature.

Broadly speaking the properties of other ferromagnetic crystalline minerals, such as hematite and magnetite, exhibit the same general features as pyrrhotite. The other minerals have not been examined so thoroughly as pyrrhotite, and there are important differences in detail; but they all possess different magnetic properties along the different axes of symmetry, and the three minerals referred to all possess one axis of conspicuously easy magnetization. The phenomena in the direction of this axis enable some of the most important consequences of Weiss's hypothesis of molecular magnetic fields to be tested. We have seen that the hysteresis curves for magnetization in this direction are very simple compared with those of the ferromagnetic metals. In the case of the minerals there is one stable value of the intensity of magnetization I_c which is practically independent of the external field. The direction of I_c may be positive or negative, depending on the previous treatment of the specimen, but otherwise it is quite definite. The evidence for regarding the value $I_x = 0$, which is also permanent within a more limited range of treatment, as a mixture of equal amounts of $+I_c$ and $-I_c$ seems quite satisfactory: the most important argument being that there is no continuous change but a sudden jump from $I = 0$ to $I = \pm I_c$. The minerals thus possess a definite value I_c of stable intrinsic magnetization which is different from zero. This is in accordance with the requirements of the theory. The reason why the ferromagnetic metals may be expected to be less definite in this respect will be briefly referred to later.

We also saw on p. 382 that this theory led to a simple relation between the intensity of permanent magnetization and the temperature (equation (16), p. 383). This relation is the same for all substances, provided the intensity I_c is expressed in terms of the greatest possible value of I_c (the value at $T=0$) and the temperature T in terms of the temperature T_0 at which the ferromagnetism disappears: that is to say, the maximum value of I_c and the absolute temperature T_0 are to be taken as the units of intensity of magnetization and temperature respectively. This theorem has been tested experimentally by Weiss* in the case of magnetite. He finds that the theoretical curve is followed with great accuracy except in the immediate neighbourhood of $T=T_0$ and at very low temperatures. Even in these regions the deviations are not very large. The experiments extend from − 79° C. to the critical temperature $T_0=+587°$ C. When one considers the wide range of temperature covered, and the fact that there are no disposable constants in the formula, this agreement must be regarded as a remarkable confirmation of the theory.

There is one point which seems to call for further discussion at this stage. Equation (16) is derived from the equations

$$\frac{I}{I_0}=\frac{\cosh a}{\sinh a}-\frac{1}{a}, \quad a=\frac{MH}{RT} \text{ and } H=H_1+H_2$$

by putting the external field H_1 equal to zero and $H_2=\lambda I$, the derived equation (15) being used to eliminate the constants. Since I is a continuous function of T when $H_1=0$ it should also be a continuous function of H_1 and T when H_1 is allowed to vary. At first sight this appears to be contradicted by the experimental results. For, so far as the experiments have shown, there is no appreciable change in the value of $I=I_c$ as the external field H_1 is increased from zero to the highest values available in the laboratory. This would seem to be a fatal objection to the theory unless the values of the internal field λI were so great that the largest values of H_1 attainable were negligible in comparison. In that case the behaviour would be sensibly the same as for $H_1=0$ even in the highest fields which can be obtained. We shall see in the next paragraph that it is possible to deduce the values of the coefficients λ by an independent method, from experimental data. The values of λ so obtained show that λI *is* extremely large compared with the magnetic forces at our disposal; so that this objection falls to the ground.

---

* *Journal de Physique*, vol. VI. p. 665 (1907).

Properties near the Critical Temperature.

In this neighbourhood we have seen that

$$\frac{I}{I_0} = \frac{a}{3}, \quad a = \frac{M}{RT}(H_1 + H_2), \quad H_2 = \lambda I.$$

Thus
$$\frac{I}{I_0} = \frac{1}{3}\frac{M}{RT}(H_1 + \lambda I).$$

The critical temperature is $T_0 = \frac{\lambda M I_0}{3R}$. Thus

$$\frac{I}{I_0}\left(1 - \frac{T_0}{T}\right) = \frac{H_1}{\lambda I_0}\frac{T_0}{T} \quad \text{or} \quad (T - T_0) I = \frac{H_1 T_0}{\lambda} \quad \text{......(18).}$$

It follows that at temperatures sufficiently near to T_0 the product of the intensity of magnetization I, and the difference $T - T_0$ between the actual and the critical temperature, will remain constant when the magnetizing force H_1 is maintained constant. The I, T curves will be portions of rectangular hyperbolas whose

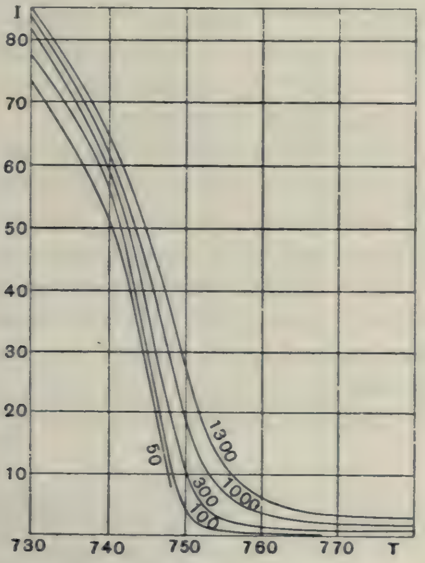


Fig. 49.

parameter is proportional to the field strength H. That this is at least approximately true is proved by the accompanying diagram, which represents the results of measurements by P. Curie on a specimen of iron in the neighbourhood of the critical temperature. The values of H are written alongside the corresponding curves.

It is evident that equation (18) may be used to determine the values of the coefficient λ, since T, T_0, I and H_1 are given by the experiments. From measurements of this kind the following values of λ and H_2 have been deduced by Weiss:

Iron $\lambda = 3{,}850$ $H_2 = 6{\cdot}56 \times 10^6$ E.M.U.
Nickel $\lambda = 12{,}700$ $H_2 = 6{\cdot}35 \times 10^6$ „
Magnetite $\lambda = 33{,}200$ $H_2 = 14{\cdot}3 \times 10^6$ „

The values of H_2 are enormous compared with the magnetic fields which can be obtained in the laboratory, so that the peculiar result that I_c does not appear to vary with the external field is accounted for satisfactorily.

Abrupt Changes in Magnetic Properties.

At high temperatures iron exhibits a number of more or less abrupt changes in its magnetic properties. Below 756° C. it exhibits the characteristic ferromagnetic properties usually associated with the metal. Between 820° C. and 920° C. it appears to be incapable of permanent magnetization, but it exhibits the rapid diminution of susceptibility with rising temperature which, as we have seen, should characterize ferromagnetic substances in the neighbourhood of the critical temperature. Between 920° C. and 1280° C. it behaves like a typical paramagnetic substance, the susceptibility varying as the inverse absolute temperature. Above 1280° C. it has similar properties, except that there is a sudden increase in the susceptibility at this temperature. This varying behaviour has been attributed to the existence of different modifications of iron within each of the limits of temperature specified. These modifications are denoted by Fe α, Fe β, Fe γ and Fe δ.

It has been pointed out by Weiss[*] that the constants which determine the magnetic behaviour of these different forms of iron exhibit simple numerical relations. These may be determined as follows:

---

[*] *Journal de Physique*, vol. VI. p. 685 (1907).

Fe α. In the neighbourhood of T_0 when I and a tend towards zero

$$I = I_0 \frac{a}{3} = \frac{I_0 MH}{3RT} = \frac{I_0 . NM . H}{3NRT},$$

where N is the number of magnetic molecules per unit volume of the material. But $NM = I_0$ and NRT is the pressure p which would be exerted by the substance if it were gaseous and occupied the same volume at the same temperature. Thus if δ is the density of the substance the constant

$$\chi = \frac{I}{H\delta} = \frac{I_0^2}{3p\delta}.$$

This formula is derived on the supposition that each kinetic molecule forms one magnetic molecule. If however n of the latter go to make up one of the former the right-hand side will have to be multiplied by n. Putting in known values of I_0 and δ, and the value of p derived from obvious data, one finds that

$$C = \chi T = 0{\cdot}00165n.$$

Fe β. We may attribute the quasi-ferromagnetic behaviour of this body to the fact that the external field H_1 is helped by the molecular field λI. Thus if $\chi\delta$ is the true susceptibility (δ = density) and $C\delta$ is the true value of Curie's constant

$$\frac{T}{C} = \frac{1}{\chi} = \delta \frac{H_1 + \lambda I}{I} = \frac{1}{\chi'} + \lambda\delta,$$

where $\chi'\delta$ is the measured value of the susceptibility. Comparing this with the equation

$$I(T - T_0) = \frac{H_1 T_0}{\lambda},$$

we see that, since H_1 is arbitrary,

$$C = T_0/\lambda\delta = \chi'(T - T_0).$$

Substituting the experimental value of $\chi' = 5{\cdot}13 \times 10^{-4}$ at 820° C. one finds

$$C = 0{\cdot}00164 \times 2.$$

Fe γ. The experimental values of the susceptibility of this paramagnetic body give the following values for C. At 940° C., $C = 0{\cdot}00172 \times 2$, and at 1280° C., $C = 0{\cdot}00182 \times 2$.

Fe δ. In a similar way the value of C for this paramagnetic body is found to be $0{\cdot}00198 \times 3$ at 1280° C. and $0{\cdot}00173 \times 3$ at 1336° C.

The interesting point is that all the values of C above are equal, within the limits of experimental error, either to twice or to three times the common factor 0·0017.

The occurrence of sudden changes in the magnetic qualities is by no means confined to iron. Similar features are presented by the other ferromagnetic substances which have been examined. In fact abrupt changes also occur even with diamagnetic substances. Thus Curie found that there was a large drop in the value of the diamagnetic constant of bismuth when fusion occurred. Moreover the diamagnetic constant of the molten bismuth was independent of the temperature, whereas this was not the case with solid bismuth. In the case of tin, which is sometimes diamagnetic and at others paramagnetic, according to the temperature, a number of transition points have been observed*. These facts support the view that a considerable part at least of the magnetic properties of bodies are determined by the occurrence of systems of considerable size rather than the atoms or sub-atomic structures†. It is of interest to form an estimate of the local strength of the molecular fields on the hypothesis that the apparent saturation of iron is due to the equilibrium between the internal fields and the kinetic energy of thermal agitation. Weiss (*loc. cit.* p. 688) estimates that for iron at ordinary temperatures $I/I_0 = 0{\cdot}91$ about, whence, making use of Langevin's formula, $a = 11{\cdot}3$. Thus

$$\frac{MH}{RT} = \frac{NM \,.\, H}{NRT} = \frac{NM \,.\, H}{p} = 11{\cdot}3,$$

where p is the pressure exerted by the magnetic molecules, supposed gaseous, and filling the same space as the metal. Putting $NM = 2000$ gauss and $p = 2 \times 10^9$ dynes per sq. cm. one finds $H =$ about 11×10^6 lines per sq. cm. This rough calculation agrees as to order of magnitude with the more accurate estimate which was obtained on p. 390, and supports the conclusion to which we have already been driven that the behaviour of simple ferromagnetic substances, in the fields available in the laboratory, will be practically the same as when $H_1 = 0$.

---

* Du Bois and Honda, *Versl. Kon. Ak. van Wetensch. Amsterdam*, vol. XII. p. 596 (1910).

† Cf. Oxley, *Camb. Phil. Proc.* vol. XVI. p. 486 (1912).

Other Properties of the Ferromagnetic Metals.

Whatever the ultimate explanation of the very interesting properties of ferromagnetic crystals like pyrrhotite and magnetite may be, it is probable that they furnish an indication of the direction in which we should look for an explanation of the behaviour of ferromagnetic metals. It is well known now that all metals are complex aggregates of minute crystals. When the metals are impure, as is the case with most specimens of iron, for example, the crystals may vary considerably in composition as well as size. It is therefore reasonable to expect the behaviour of iron to resemble that of an irregular matrix of small crystals of, let us say, pyrrhotite. The behaviour of such a matrix can readily be calculated. If, for the moment, we neglect the small susceptibility, parallel to the y and z axes, of any crystal selected at random, the latter will not develop any magnetization until the x component of H reaches the critical value H_c. Thus all the crystals, supposed initially in random azimuths, will not become magnetized simultaneously, and the intensity of magnetization will not approach saturation suddenly, as with a single crystal, but gradually. This agrees, of course, with the behaviour of iron. On the other hand there would on this view be no magnetization until $H = H_c$, whereas iron has a definite susceptibility for $H = 0$. This can be accounted for when the small magnetizations parallel to Oy and Oz are considered, and by considering such additional factors as the lack of homogeneity of the material and the possibility of local inequalities in the magnetization, it is probable that the behaviour of any particular specimen of iron could be imitated exactly by a model of this kind. For these reasons the study of ferromagnetic crystals is probably of fundamental importance towards the understanding of ferromagnetic materials in general. It is worthy of remark that in the case of iron deposited electrolytically in a magnetic field, hysteresis curves have been obtained which are almost rectangular like those given by pyrrhotite in the magnetic plane*.

---

* Maurain, *Journal de Physique*, Ser. 3, vol. x. p. 123 (1901).

The Ultimate Magnetic Elements.

In the case of magnetite above the critical temperature, Weiss* finds that the value of χ is not proportional to T^{-1} as in the case of the typical paramagnetic substances, but that the graph of χ^{-1} against T consists of a series of straight lines at different inclinations. Since $C = mI_1^2/3R$, where C is Curie's constant per unit mass, m is the molecular weight, and I_1 the saturation intensity of magnetization per unit mass, this result might be interpreted as arising from a variation of either m or I_1, or both. Weiss rejects the variation of m on the ground that it does not lead to simple results. On the other hand, if m is assumed to be constant, the resulting values of I_1 calculated from the experimental values of C are to each other, within the limits of experimental error, in the ratio of the numbers 4, 5, 6, 8 and 10. To explain this Weiss is led to make the hypothesis that the magnetic properties of substances arise from the presence of an ultimate unit, the "magneton," in the atoms of the substance. It is apparently necessary that these elements should be capable in some way of annihilating each other temporarily, as the same substance may contain different numbers of magnetons at different temperatures.

In the case of other substances the number of these elements per gramme atom or molecule may be determined in different ways. In the case of ferromagnetic substances it may be deduced from the saturation intensity of magnetization at the absolute zero. In the case of solutions of paramagnetic substances all that is required is the value of the constant $C = mI_1^2/3R$. In these ways values of this number have been deduced for nickel, cobalt, iron and a large number of salts of iron, cobalt, manganese, chromium, copper, uranium, vanadium and the rare earths. The integral numbers vary from 4 to 56 in different cases. In the case of two salts of vanadium there is no indication of an approach to simple integral multiples; but the agreement in a great many cases in which the integral number is small is so exact that there seems little room for doubt that the idea of the magneton rests on a substantial basis of fact.

It is worth while to add that a theory of the behaviour of the atom which has had very considerable success in some other branches of physics leads to the existence of a magneton. This theory supposes the atoms to be made up of rings of electrons, revolving round a positive kernel of small dimensions, having a charge equal and opposite to the sum of the charges on the electrons. The steady motions are subject to dynamical laws and do not give rise to radiation in appreciable quantities. Radiation occurs when the electrons move from one stable configuration to another, and its frequency ν is determined in accordance with the quantum hypothesis by the equation $h\nu = W$, where h is Planck's constant, and W is the change in energy which accompanies the change in the configuration. Nicholson* and Bohr† have shown that under these circumstances the value of T/ν, where T is the kinetic energy of the electron and ν its orbital frequency, is always an integral multiple of $h/2$. Now we have seen that the average magnetic moment of the magnet, which is equivalent to n electrons revolving in a circular or elliptic orbit, is $M = neS/\tau$, where S is the area of the orbit and τ the time of description. Thus if m is the mass of an electron

---

* *Le Radium*, vol. VIII. p. 301 (1911).

$$M = \frac{n}{2\pi} \frac{e}{m} \frac{T}{\nu}.$$

Thus it follows from this type of theory that M will always occur in integral multiples of M_0, where

$$M_0 = \frac{e}{m} \frac{h}{4\pi}.$$

Putting $e/m = 1{\cdot}77 \times 10^7$ E.M.U. and $h = 6{\cdot}55$ erg sec., this gives

$$M_0 = 9{\cdot}23 \times 10^{-21}.$$

This is nearly six‡ times as large as the value of the magneton found by Weiss from experimental considerations. The experimental value of the magneton is $1{\cdot}64 \times 10^{-21}$ in the same units.

Mechanical Reaction caused by Magnetization.

On the theory of magnetism which we have been discussing it appears to the writer§ that we might expect to observe a rotational mechanical reaction when a bar of iron is magnetized.

---

* *Monthly Not. Roy. Astr. Soc.* vol. LXXII. p. 679 (1912).

† *Phil. Mag.* vol. XXVI. p. 1 (1913).

‡ Since this was written I have learnt from a conversation (July 1913) with Dr Bohr, who had made similar calculations, that a more exact experimental value of the magneton makes this ratio exactly five.

§ *Phys. Rev.* vol. XXVI. p. 248 (1908).

For simplicity suppose that the magnetization arises entirely from the orbital motion of negative electrons whose charge is e. Let there be N of them per unit volume, and let A_z denote the average value of the projections of the areas of the orbits perpendicular to the axis of magnetization, divided by the times in which they are described. Then I_z, the magnetic moment per unit volume, is given by $I_z = NeA_z$. The resultant magnetization is taken to be parallel to the axis of z.

We shall now calculate the moment of momentum of the revolving electrons about the z axis. Consider any approximately circular orbit, the coordinates of whose centre are given by x_0, y_0, z_0. Let the coordinates of the revolving electron referred to this centre at any instant be ξ, η, ζ. The moment of momentum of the electron about the z axis is then

$$m\left\{(x_0+\xi)\frac{\partial\eta}{\partial t}-(y_0+\eta)\frac{\partial\xi}{\partial t}\right\}=m\left(x_0\frac{\partial\eta}{\partial t}-y_0\frac{\partial\xi}{\partial t}+\xi\frac{\partial\eta}{\partial t}-\eta\frac{\partial\xi}{\partial t}\right).$$

Averaging this over a complete revolution the mean values are

$$\overline{x_0\frac{\partial\eta}{\partial t}}=0=\overline{y_0\frac{\partial\xi}{\partial t}},$$

and $$\overline{\xi\frac{\partial\eta}{\partial t}}=-\overline{\eta\frac{\partial\xi}{\partial t}}=\frac{\text{area of projected orbit}}{t}=A_z.$$

Hence the average moment of momentum about any axis is independent of the position of that axis so long as its direction is the same. It is equal to $2mA_z$. The moment of momentum U_z per unit volume is

$$U_z = 2NmA_z = 2\frac{m}{e}I_z \quad \text{..................(19).}$$

It is thus equal to the intensity of magnetization multiplied by $2m/e$.

By the principle of the conservation of momentum the moment of momentum thus created must be balanced by an equal moment about the same axis. This reaction might conceivably occur either (1) on the electromagnetic system producing the exciting field, or (2) on the atoms of the magnetic material. In the former case the effect should depend simply on I_z and not involve e/m. In the latter case the effects observed depend on the looseness of the atoms. If they were free to rotate without affecting the neighbouring atoms, the moment of momentum of the orbits might be compensated by local rotations of the non-magnetic matter. In that case the magnetic material would not have any tendency to turn as a whole. On the other hand the fact that the magnetic properties of the material are very susceptible to changes of temperature makes it very unlikely that the connection between the magnetic atoms and the neighbouring matter is a loose one. It seems therefore most probable that the moment of momentum created in this way will be compensated by a motion of the magnetic material as a whole. Experiments which have been made to detect this effect have not led to a decision as to whether it exists or not.

Specific Heats of Ferromagnetic Substances.

If a substance is magnetized to an intensity I in a field of strength H, and I is proportional to H, the energy of the system is changed by an amount $\frac{1}{2}HI$. On account of the very large magnitude of the internal fields this energy is comparable with the whole thermal energy in the case of the ferromagnetic substances. This is true even when the average intensity of magnetization is zero. As the energy of the molecular fields is a function of the temperature, a very considerable part of the specific heat may arise in this way. The molecular magnetization diminishes with rising temperature so that additional energy must be supplied in order to overcome the attraction of the elementary magnets. The effect thus involves an increase in the specific heat of the substance. Since $H_2 = \lambda I$ this part of the specific heat is $\frac{\lambda I}{J}\frac{\partial I}{\partial T}$, where J is the mechanical equivalent of heat. Since I disappears suddenly at the critical temperature so will this additional specific heat. Weiss and Beck* have shown that the part contributed by the internal magnetic energy in this way accounts quantitatively for the anomalous specific heats of ferromagnetic substances.

---

* *Journal de Physique*, Ser. 4, vol. VII. p. 249 (1908).

CHAPTER XVII

THE KINETIC THEORY OF ELECTRONIC CONDUCTION

Thermodynamics and the Kinetic Theory of Matter.

(i) *Entropy and Probability.*

When a material system is isolated in such a way that its total energy U and total volume v remain constant, its physical state will nevertheless, in general, change with lapse of time. This is evident since the two variables U and v do not completely specify the condition of the system—if we understand by the condition of the system the way in which it reacts to instruments which measure such quantities as pressure, volume and temperature, which characterize matter in bulk rather than the individual molecules of which we believe it to be composed. For example, we might have two systems having the same material composition and the same values of U and v, but the temperatures of corresponding points of the two systems might be different. The two systems if left alone would then change in different ways as time elapsed. The changes which ensue are not capricious but definite. So far as physical measuring instruments are sensible of them, they tend to the establishment of a definite end condition. The final state is characterized by the fact that a certain function called the entropy (S) of the system has attained the maximum value consistent with the imposed conditions.

The entropy may be considered to be defined by the differential equation

$$dS = \frac{1}{T}(dU + pdv) \quad \ldots\ldots\ldots\ldots\ldots\ldots(1),$$

where T is temperature and p pressure. From this definition and the second law of thermodynamics, it may be shown that any reversible change occurring in an isolated system leaves the value of S unaltered, whereas any irreversible change increases S. Thus the final state is that for which S is a maximum. It may also be shown that the change in S produced by any reversible action depends only on the initial and final states of the system and not at all on the way in which the change has been effected, provided it is reversible throughout. Thus S is a perfect differential with respect to the independent variables in terms of which the state of the system is described.

The final state is only steady as regards quantities like pressure, temperature and so on, which are usually taken to be sufficient to describe the behaviour of matter in bulk. If we made use of instruments fine enough to determine the motions of the individual molecules there is every reason to believe that the system would be found to be the seat of very lively, never ceasing changes. The final steady state is therefore one of statistical equilibrium merely. The actual spatial distribution of the individual molecules and the distribution of the momentum and velocity which they possess are both constantly changing. On the average, however, and actually if the system contains an indefinitely large number of molecules, the distribution of the molecules in space and of momentum and velocity among the molecules is definite. We shall now attempt to discover what this law of distribution is.

The number of ways in which a given amount of energy may be distributed among an indefinitely large number of molecules is clearly infinite to a very high order. Some of these are much more probable than others and there will be one distribution which is the most probable. Jeans* has shown that in the statistically steady state which is independent of the time the most probable distribution is infinitely probable compared with the others; so that if we can find the most probable distribution we shall have obtained the actual distribution for all practical purposes.

Boltzmann pointed out the intimate connection between the probability of a given state of a system and the entropy of the system. In the final state of an isolated system we have seen that both the entropy and the probability of the system have attained a maximum value. In the intermediate stages both quantities are moving towards the maximum. Let us now introduce the hypothesis that the entropy of a system is a function of its probability only. This course is evidently permissible, since we have not yet defined the probability of a system precisely. We shall have to take care, when we come to do this, that our definition does not conflict with the hypothesis. Let us see what conclusions we may arrive at from the general conceptions of entropy and probability with the aid of our hypothesis.

---

* *Dynamical Theory of Gases*, Chap. III.

Consider two entirely separate material systems, let us say two stars so far apart that the interaction of their radiations may be disregarded. Then we have, by hypothesis, $S_1 = f(w_1)$ and $S_2 = f(w_2)$, where S_1 and S_2 are the entropies and w_1 and w_2 the probabilities of the systems separately. If S and w are respectively the entropy and probability of the two systems considered together, we have

$$f(w) = S = S_1 + S_2 = f(w_1) + f(w_2).$$

But

$$w = w_1 w_2,$$

hence

$$f(w_1 w_2) = f(w_1) + f(w_2),$$

so that

$$S = k \log w \quad \text{.......................(2)},$$

where k is a universal constant. Thus the difference between entropy and probability is only that one combines by addition and the other by multiplication.

Of the total number of molecules under consideration let us suppose that in the steady state the fraction

$$f(x, y, z, p, q, r)\, dx\, dy\, dz\, dp\, dq\, dr \quad \text{...........(3)}$$

have their centres between x and $x + dx$, y and $y + dy$, and z and $z + dz$, and their momenta between the components p and $p + dp$, q and $q + dq$, and r and $r + dr$. We shall suppose that the six variables x, y, z, p, q and r completely describe the state of the particles. This is equivalent to treating them as massive points subject to the action of forces, and although not general enough for many problems in dealing with gases, is sufficiently so for the electrical problems in which the particles under consideration are electrons. Another interpretation which may be given to $f(\quad)$ is that it is the probability that a particle selected at random should have its six coordinates within the assigned limits. We shall suppose that f is a continuous function which can be differentiated, and proceed to determine the probability of the state of the gas which corresponds to any assigned form of the function f.

We may consider the gas to be represented by a series of points in a six-dimensional space, lengths measured along the axes of which give the values of x, y, z and p, q, r, the components of the distance of each particle from a fixed centre and of its resultant momentum respectively. Each particle is represented by one point, and if we know the density of such points at every part of the six-dimensional diagram we shall have a complete picture of the state of the gas. Now divide up the whole of the space so that the six-dimensional elements of volume

$$dxdydzdpdq\,dr = d\sigma$$

are everywhere equal. We shall define the probability of a given distribution f as the number of ways in which the given distribution may be constructed by distributing the total number n of particles among the different elements of volume. We shall consider the elements $d\sigma$ to be so small that the state of the particles which are defined by the limits of $d\sigma$ is to be considered precisely the same for all of them. This would introduce a difficulty if we considered the diagram to represent an actual instantaneous state of the system. The difficulty can be overcome by considering it to represent a large number of successive states. In the former case the number of particles in each element of volume is necessarily limited; in the latter case it may be made as large as we please by contemplating a sufficiently large number of successive states. Since f is given, the number of particles in the element $d\sigma$ is $nf\,d\sigma$. Since the particles in any element are to be treated as alike, any rearrangement of them within the element will not give rise to a fresh distribution. The problem therefore is to find the number of ways in which n like objects may be distributed among the totality of the compartments $d\sigma$ which make up the whole of the space, when the same distribution is considered to arise wherever the same particles occur in the same elements of volume, no matter how they are arranged within the element of volume. The number of ways is clearly

$$n! \div \Pi\,(nf\,d\sigma)! \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4),$$

where Π denotes the continued product taken over all the elements. As an example we may consider the number of ways in which all the particles may be given the same position and momentum. This is evidently equal to one, which is also the value given by (4), since, for all the elements $d\sigma$ except the one in which the particles lie, we have $(nf\,d\sigma)! = 0! = 1$, and for the remaining element

$$(nf\,d\sigma)! = n!.$$

Since the size of the elements $d\sigma$ is arbitrary the expression for the probability above will only be the value in terms of some arbitrary standard. It is not necessary, however, for us to determine the value of the standard, since it is possible to arrive at results which are independent of $d\sigma$ without doing so.

By combining (4) with (2) we have

$$S = k \log n! - k \Sigma \log (nf\,d\sigma)! \quad \ldots\ldots\ldots\ldots\ldots(5).$$

Now we can always make $nf\,d\sigma$ as large as we please by taking n big enough. When N is any very large number we have Stirling's well-known approximation

$$\left.\begin{aligned} N! &= \left(\frac{N}{e}\right)^N \sqrt{2\pi N} \\ \text{or} \qquad \log N! &= N(\log N - 1) + \tfrac{1}{2}\log 2\pi N \\ &= N(\log N - 1) \end{aligned}\right\} \ldots\ldots\ldots\ldots(6)$$

with sufficient approximation, since we may neglect $\log N$ compared with N when N is large. Thus (5) may be written

$$\begin{aligned} S &= k \log n! - k \Sigma\, nf\,d\sigma (\log nf\,d\sigma - 1) \\ &= k \log n! - k \Sigma\, nf\,d\sigma (\log nf + \log d\sigma - 1) \\ &= k [\log n! - n(\log d\sigma - 1)] - k \Sigma\, nf\,d\sigma \,.\, \log nf, \end{aligned}$$

since all the $d\sigma$'s are equal and $\Sigma\, nf\,d\sigma = n$. Since n is constant we obtain

$$S = \text{const.} - k \int nf \log nf\,d\sigma \quad \ldots\ldots\ldots\ldots\ldots(7).$$

If the particles under consideration make up the whole of the system then (7) will be the complete expression for the entropy. Without any more elaborate analysis we may add to S a part which is independent of the particles under consideration and therefore independent of f. With this understanding we may put

$$S = S_0 - k \int nf \log nf\,d\sigma \quad \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

where S_0 is made up of the entropy of the foreign parts of the system together with an arbitrary constant.

(ii) *The Law of Distribution.*

Now if the total energy and volume of the system are constant, the final steady state is characterized by a maximum value of S. This fact is sufficient to determine the function nf. We shall suppose the energy U of the system to consist of three parts, (1) a part U_0 which is independent of the n molecules, (2) the potential energy W of the n molecules, and (3) their kinetic energy L; i.e.

$$U = U_0 + W + L \quad \text{.....................(9)},$$

where
$$W = \int \phi(x, y, z) f \, d\sigma = \int W_\tau f \, d\sigma \quad \text{...............(10)},$$

and
$$L = \frac{1}{2m} \int (p^2 + q^2 + r^2) f d\sigma = \int L_\sigma f \, d\sigma \quad \text{.........(11)}.$$

Thus W_τ and L_σ are independent of f. Since the entropy is to be a maximum for the actual function f we have by varying nf

$$\delta S = \int (\log nf + 1) \, \delta(nf) \, d\sigma = 0 \quad \text{.........(12)};$$

also since the total number n of particles is constant

$$0 = \int \delta(nf) \, d\sigma \quad \text{.....................(13)};$$

finally since the total energy of the system is constant

$$0 = \int (W_\tau + L_\sigma) \, \delta(nf) \, d\sigma \quad \text{...............(14)}.$$

These equations are satisfied for all possible variations of nf if

$$\log nf = -k_0 (W_\tau + L_\sigma) + \text{const.},$$

where k_0 is constant throughout the system. Thus

$$nf \, dxdydzdpdqdr = Ae^{-k_0 (W_\tau + L_\sigma)} \, dxdydzdpdqdr \text{...(15)},$$

where A is constant throughout the system.

We may determine k_0 from a knowledge of the mean kinetic energy L/n of the particles. We have evidently

$$\frac{L}{n}=\frac{\iiiint\!\!\iint \frac{p^2+q^2+r^2}{2m} A e^{-k_0(W_\tau+L_\sigma)}\,dx\,dy\,dz\,dp\,dq\,dr}{\iiiint\!\!\iint A e^{-k_0(W_\tau+L_\sigma)}\,dx\,dy\,dz\,dp\,dq\,dr}$$

$$=\frac{1}{2m}\frac{\iiint (p^2+q^2+r^2)\,e^{-\frac{k_0}{2m}(p^2+q^2+r^2)}\,dp\,dq\,dr}{\iiint e^{-\frac{k_0}{2m}(p^2+q^2+r^2)}\,dp\,dq\,dr}$$

$$=\frac{3}{2}\frac{1}{k_0} \text{ or } k_0=\frac{3}{2}\frac{n}{L} \qquad\text{(16).}$$

Moreover since

$$n=A\iiiint\!\!\iint e^{-k_0(W_\tau+L_\sigma)}\,dx\,dy\,dz\,dp\,dq\,dr$$

$$=A\iiint_\tau e^{-k_0W_\tau}\,d\tau\times\iiint_{-\infty}^{+\infty} e^{-\frac{k_0}{2m}(p^2+q^2+r^2)}\,dp\,dq\,dr,$$

$$A=\left(\frac{3n}{4\pi mL}\right)^{\frac{3}{2}}\frac{n}{\iiint e^{-k_0W_\tau}\,d\tau} \qquad\text{(17).}$$

From equation (15) we see that if $\nu_\tau\, d\tau$ is the number of particles, with no restriction as to momentum or velocity, in the element of volume $d\tau = dx\,dy\,dz$, then

$$\left.\begin{aligned}\nu_\tau d\tau &= A e^{-k_0W_\tau}\,d\tau\iiint_{-\infty}^{\infty} e^{-k_0\frac{p^2+q^2+r^2}{2m}}\,dp\,dq\,dr\\ &=\frac{ne^{-k_0W_\tau}\,d\tau}{\iiint e^{-k_0W_\tau}\,d\tau}\end{aligned}\right\}\ldots\text{(18).}$$

Thus if we compare any two different elements of volume $d\tau$ and $d\tau'$

$$\log\frac{\nu_\tau}{\nu_{\tau'}}=k_0\,(W_\tau'-W_\tau) \qquad\text{(19).}$$

If we denote the probability that a particle situated in the element of volume $d\tau$ has components of momentum which lie between p and $p+dp$, q and $q+dq$, and r and $r+dr$, by

$$F(p, q, r)\,dp\,dq\,dr,$$

then

$$nf\,dx\,dy\,dz\,dp\,dq\,dr=\nu_\tau\,d\tau\,F\,dp\,dq\,dr=Ae^{-k_0(W_\tau+L_\sigma)}\,d\tau\,dp\,dq\,dr,$$

and from (18)

$$F(p, q, r)\,dp\,dq\,dr = \frac{e^{-k_0\frac{p^2+q^2+r^2}{2m}}\,dp\,dq\,dr}{\iiint_{-\infty}^{\infty} e^{-k_0\frac{p^2+q^2+r^2}{2m}}\,dp\,dq\,dr}$$

$$= \left(\frac{k_0}{2m\pi}\right)^{\frac{3}{2}} e^{-k_0 L_\sigma}\,dp\,dq\,dr \quad \text{..................(20)}$$

$$= \left(\frac{3n}{4\pi m L}\right)^{\frac{3}{2}} e^{-k_0\frac{p^2+q^2+r^2}{2m}}\,dp\,dq\,dr \quad \text{.........(21).}$$

It follows from equations (20) and (21) that F is entirely independent of the potential energy W_τ of the particles in the element $d\tau$. The distribution of velocity among the particles, and also of kinetic energy, is entirely independent of their potential energy. We see from (19) that the variation of the potential energy causes the numbers of particles in equal volumes to vary from point to point, although their mean kinetic energy is the same everywhere.

An apt illustration is furnished by the equilibrium of a column of gas, in an enclosure at constant temperature, subject to the action of gravitation. The mean kinetic energy of the particles is the same at every point, as also is the way in which the kinetic energy is distributed among the particles. On the other hand, the density of the gas is greatest at the bottom of the vessel where the gravitational potential energy is least. The law of variation of density with height is readily deduced from equation (19).

The pressure p exerted by the particles at a point where the number in unit volume is ν is easily calculated. Consider a smooth impenetrable surface, whose plane is perpendicular to the axis of x, to be placed at the point. On impact with the surface each colliding particle will have its x component of momentum reversed. The momentum communicated to unit area of the surface in unit time is evidently

$$p = 2\nu \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{p^2}{m}\left(\frac{k_0}{2m\pi}\right)^{\frac{3}{2}} e^{-k_0\frac{p^2+q^2+r^2}{2m}}\,dp\,dq\,dr$$

$$= \frac{\nu}{k_0} = \frac{2}{3}\nu\frac{L}{n} \quad \text{..(22).}$$

Since we are neglecting effects which depend upon the size and mutual potential energy of the particles they will behave like a perfect gas and satisfy the equation $pv = R_1 T$. If R is the value of R_1 reckoned for an amount of gas equal to one molecule, this may be written

$$p = \nu RT \qquad \text{.........................(23).}$$

By comparison with (22) we see that

$$k_0 = 1/RT \qquad \text{.........................(24),}$$

and the average kinetic energy $\frac{L}{n} = \frac{3}{2} RT$. We shall frequently make use of (19) in the form

$$\log \frac{\nu}{\nu'} = \frac{W' - W}{RT} \qquad \text{.................(19 } a\text{).}$$

(iii) *The Constant k.*

Having determined the nature of the function nf we can now write the expression (8) for the entropy of the system in the form

$$S = S_0 + k \int A \left(k_0 \{ W_\tau + L_\sigma \} - \log A \right) e^{-k_0 (W_\tau + L_\sigma)} \, dx\, dy\, dz\, dp\, dq\, dr \qquad \text{......(25).}$$

To determine the universal constant k we need only consider the simplest case, that of a monatomic gas for which W_τ is everywhere zero. Equation (25) then becomes

$$\left.\begin{aligned} S &= S_0 + kV \int_{-\infty}^{\infty} A(k_0 L_\sigma - \log A)\, e^{-k_0 (W_\tau + L_\sigma)} \, dp\, dq\, dr \\ &= S_0 + kn \left[\frac{3}{2} \left(1 - \log \frac{3}{4\pi m} + \log \frac{L}{n} \right) + \log \frac{V}{n} \right] \\ &= S_0' + kn \left[\frac{3}{2} \log L + \log V \right] \end{aligned}\right\} \text{.... (26),}$$

where S_0' is a new constant, which is independent of L and V. In the present case L is the total energy of the system, since the potential energy is zero; so that

$$dS = \frac{dL + p\,dV}{T},$$

whence

$$\left(\frac{\partial S}{\partial L}\right)_V = \frac{1}{T} \text{ and } \left(\frac{\partial S}{\partial V}\right)_L = \frac{p}{T} \qquad \text{..............(27).}$$

On the other hand we obtain from (26)

$$\left(\frac{\partial S}{\partial L}\right)_V = \frac{3}{2}\frac{kn}{L} \text{ and } \left(\frac{\partial S}{\partial V}\right)_L = \frac{kn}{V},$$

whence, by comparison with (27),

$$pV = knT$$

and by comparing this with the equation of a perfect gas we have

$$k = R = (Tk_0)^{-1}.$$

The numerical value of R is known to the same degree of precision as that to which we know the value of the mass of an atom. With the accuracy which has recently been reached in this branch of physical measurement the different methods of determining R are consistent to within about 1 °/ₒ. One of the best values, deduced by Planck (see Chap. XV, p. 356) from the constants in the complete radiation formula, is

$$R = k = 1{\cdot}346 \times 10^{-16} \frac{\text{gm. cm.}^2}{\text{sec.}^2 \text{ deg.}}.$$

The Theory of Metallic Conduction.

The view, that the transportation of electricity in metallic conductors, like the corresponding phenomenon exhibited by electrolytes, is due to the motion of minute charged particles, was a definite feature of the old Weberian formulation of electrodynamics. The modern development is due to the labours of Riecke, Drude, J. J. Thomson and others. According to the most prevalent form of the theory, and the one which we shall adopt, the atoms of the metal are regarded as continually dissociating into a negative electron and a positively charged residue. It follows from the kinetic theory considerations which we have just discussed that the electrons will be moving in all directions in the interior of the metal with very high speeds. They will in fact possess the same distribution of velocity as would an uncharged molecule of equal mass, and their mean kinetic energy will be proportional to the absolute temperature. We should rather expect that the atoms and positively charged residues would have an equal distribution of kinetic energy, but it is possible that they are jammed in some way which prevents this coming about. Whether the atoms and positive residues are able to move or not, they are to be regarded as only oscillating about fixed positions and not travelling from point to point of the material.

We now suppose that the effect of applying an electromotive intensity is to superpose on the haphazard motion of the electrons, which arises from thermal agitation, a velocity of drift which, since they are negatively charged, is in the opposite direction to the applied force. This motion constitutes the electric current, which is thus carried entirely by the negative electrons. *A priori*, one might be tempted to suppose that the positive residues would also drift along under the influence of the electric field. If any such effect exists it must be extremely small. For if it were appreciable we should expect an electric current to transport atoms of one metal into the other across the junction between two metals. Careful experiments have been made to detect such effects, but they have always led to negative results. This objection would be removed if we supposed the positive particles to be of the same nature in different materials. There are, however, two serious objections to such a view. In the first place there is no other evidence of the liberation of such particles from atoms under conditions which can be considered at all analogous to those which hold inside conductors. In the second place the hypothesis appears to be incapable of removing such difficulties as are presented by the simpler theory.

The strongest arguments in favour of the view which asserts that the currents in metals are carried by negative electrons are as follows:

(1) Conductors when heated or illuminated are found to emit electrons into the surrounding space. The mass of the electrons thus emitted from a hot wire may, in favourable cases, be large compared with the number which, on any reasonable hypothesis, may be expected to be present in the wire at any instant, showing that electrons are continuously flowing into the wire from other parts of the system*.

(2) The typical good conductors at low temperatures are all metals: i.e. electropositive elements which are known from other phenomena to liberate electrons from their atoms readily.

---

* Cf. O. W. Richardson, *Phil. Mag.* vol. xxvi. p. 345 (1913).

(3) If the particles which carry the current did not possess an extremely small mass in proportion to their charge we should not expect to get the well-known effects produced by a magnetic field in electric currents, e.g. the change of resistance in a magnetic field and the Hall effect. These effects are unquestionably very complicated, and so far the electron theory has not been able to furnish an adequate quantitative explanation of them. On the other hand it is the only theory which has been able to account for them qualitatively.

The Simple Theory of Conduction.

In order to illustrate the problem presented by electric and thermal conduction in metallic conductors we shall consider the behaviour of the electrons to be much simpler than it is likely to be in any real case. A more exact treatment will be given later. For the present we shall make the following assumptions: (1) that the electrons all move freely for the same distance λ between two collisions, (2) that they all have the same velocity of thermal agitation v and (3) that the motion of an electron subsequent to a collision is entirely independent of its history previous to the collision. We shall also assume that the only force acting on the electrons throughout the free path λ is the applied electric intensity X. It follows from assumptions (1) and (2) that the free time $t = \lambda/v$ is the same for all the electrons. To be consistent with the requirements of the kinetic theory of matter we take the kinetic energy, $\frac{1}{2}mv^2$, of the electrons to be equal to the mean value $\frac{3}{2}RT$ of the same quantity for the atoms of a monatomic gas at the same temperature. Subject to these simplified assumptions the electrical conductivity of a metal may be calculated as follows:

If e is the electric charge of an electron, the force acting on it during its free path is Xe and its acceleration is Xe/m. If the component of the velocity of the electron, parallel to the electric intensity, at the beginning of its free path is u, the value of this component at the end of the path will be $u + \frac{Xe}{m}t$, where t is the free time. The average velocity in the direction of the electric field is therefore $\frac{1}{2}X\frac{e}{m}t$, since, all directions of v being equally probable, the average value of u over a large number of electrons is zero. Since $t = \lambda/v$ the average drift velocity of the electrons in the direction of the electric field may be written $\frac{1}{2}X\frac{e}{m}\frac{\lambda}{v}$. If n is the number of electrons in unit volume, the number of them which, in unit time, drift across a unit area drawn perpendicular to the direction of the electric force X will be $\frac{1}{2}nX\frac{e}{m}\frac{\lambda}{v}$. Each of these carries a charge e, so that the quantity of electricity transported across unit area in unit time or, in other words, the electric current density i, will be

$$i=\frac{n}{2}\frac{e^2}{m}\frac{\lambda}{v}X=\frac{ne^2\lambda v}{6RT}X.$$

Thus the specific electrical conductivity σ is

$$\sigma=\frac{i}{X}=\frac{ne^2\lambda v}{6RT} \quad\text{.....................}\quad (28).$$

It is a well-known result of experiment that for the pure metals, which are good conductors, σ is almost exactly inversely proportional to the absolute temperature T except at very low temperatures. Now e and R are universal constants and do not vary with temperature. Hence if our assumptions are to be compatible with the facts it is necessary that $n\lambda v$ should be independent of T. Since $v \propto T^{\frac{1}{2}}$ this requires that $n\lambda$ should be inversely proportional to $\sqrt{T}$. At present we are not in a position to say whether this variation with T is to be attributed to the variation of n or λ or of both of them.

Comparison with Thermal Conductivity.

The best conductors of electricity are also the best conductors of heat. Under the circumstances it is natural to attribute the two effects to the same cause, viz. the motion of the electrons. From this point of view the problem of the conduction of heat in a metal is the same as that in a gas having the same number of molecules in unit volume, the same free path and the same molecular weight and temperature, as the electrons in the metal. In our calculations we shall suppose that the amount of heat which is distributed by radiation and by the dynamical action of the molecules and positive residues on each other is negligible compared with that distributed by the rapidly moving electrons. The results can therefore only be expected to be true for good electrical conductors, since the worst electrical conductors have an appreciable thermal conductivity. Under these restrictions the problem of determining the thermal conductivity is mathematically identical with that of finding the thermal conductivity of the corresponding gas. In the notation we have previously employed the thermal conductivity k of the corresponding gas, which is given in any standard text-book on the kinetic theory of gases*, is

$$k = \tfrac{1}{2} n\lambda v R \qquad (29).$$

Comparing this with equation (28) we have

$$k/\sigma = 3\,\frac{R^2}{e^2}\,T \qquad (30).$$

Equation (30) is the expression of a simple and very remarkable conclusion. *The ratio of the thermal to the electrical conductivity has the same value at the same temperature for all good conductors: for different temperatures the value of the ratio is proportional to the absolute temperature.* The first part of this generalization is known as the law of Wiedemann and Franz, by whom it was announced as an experimental discovery about 1850; the law of temperature variation was discovered experimentally by L. Lorenz somewhat later.

The more recent experiments of Jaeger and Diesselhirst enable an accurate comparison with equation (30) to be carried out. They determined the values both of the ratio of the two conductivities at 18° C. and of its temperature coefficient, for a large series of metals and alloys. The values they found are exhibited in the table on p. 412.

Assuming that the charge on the electron is equal to that carried by a hydrogen atom in electrolysis the evaluation of R/e does not require any unfamiliar data. If n is the number of molecules in one cubic centimetre of a gas at 0° C. and 760 mms. pressure, then nR is the value of the constant R_1 in the equation $pv = R_1 T$, where $v = 1$ cm.^{3}, $p = 76 \times 13{\cdot}6 \times 981$ dynes/cm.^{2} and $T = 273°$ abs. On the other hand, since the molecule of hydrogen is diatomic ne is equal to the quantity of electricity which passes through a water voltameter when $\frac{1}{2}$ c.c. of H_2 is liberated at 0° C. and 760 mms. This is equal to ·4327 E.M. unit, so that the value of $3\,\frac{R^2}{e^2}\,T$ at 18° C. $= 6{\cdot}5 \times 10^{10}$. It will be observed that for the pure metals which are good conductors the experimental value of the

---

* For instance, Jeans's *Dynamical Theory of Gases*, p. 259.

Material	Ratio: Thermal Conductivity to Electrical Conductivity	Temperature Coefficient of this Ratio, per cent.
Copper, commercial	$6{\cdot}76 \times 10^{10}$ at 18° C.	—
Copper (1) pure	$6{\cdot}65 \times 10^{10}$ at 18° C.	·39
Copper (2) pure	$6{\cdot}71 \times 10^{10}$ at 18° C.	·39
Silver, pure	$6{\cdot}86 \times 10^{10}$ at 18° C.	·37
Gold (1)	$7{\cdot}27 \times 10^{10}$ at 18° C.	·36
Gold (2) pure	$7{\cdot}09 \times 10^{10}$ at 18° C.	·37
Nickel	$6{\cdot}99 \times 10^{10}$ at 18° C.	·39
Zinc (1)	$7{\cdot}05 \times 10^{10}$ at 18° C.	·38
Zinc (2) pure	$6{\cdot}72 \times 10^{10}$ at 18° C.	·38
Cadmium, pure	$7{\cdot}06 \times 10^{10}$ at 18° C.	·37
Lead, pure	$7{\cdot}15 \times 10^{10}$ at 18° C.	·40
Tin, pure	$7{\cdot}35 \times 10^{10}$ at 18° C.	·34
Aluminium	$6{\cdot}36 \times 10^{10}$ at 18° C.	·43
Platinum (1)	$7{\cdot}76 \times 10^{10}$ at 18° C.	—
Platinum (2) pure	$7{\cdot}53 \times 10^{10}$ at 18° C.	·46
Palladium	$7{\cdot}54 \times 10^{10}$ at 18° C.	·46
Iron (1)	$8{\cdot}02 \times 10^{10}$ at 18° C.	·43
Iron (2)	$8{\cdot}38 \times 10^{10}$ at 18° C.	·44
Steel	$9{\cdot}03 \times 10^{10}$ at 18° C.	·35
Bismuth	$9{\cdot}64 \times 10^{10}$ at 18° C.	·15
Constantan (60 Cu, 40 Ni)	$11{\cdot}06 \times 10^{10}$ at 18° C.	·23
Manganin (84 Cu, 4 Ni, 12 Mn)	$9{\cdot}14 \times 10^{10}$ at 18° C.	·27

ratio is very close in all cases to the theoretical value. The deviations are greater for the poorer conductors and in almost all cases are in the direction of values greater than the theoretical. This is what would happen if the part of the thermal conductivity which does not depend on the motion of the electrons were to become appreciable. Thus the deviations lie in the direction in which they would be expected to occur. The behaviour of alloys is exceptional, but so are most of their electrical properties. In fact Lord Rayleigh* has pointed out that the electrical resistance of alloys may be expected to be unduly high on account of the existence of a "false resistance" arising from the Peltier heating effect at the junction of parts of the material of varying composition. It is also desirable to mention that Lees† has found that the divergence of the values of $k/\sigma T$ for different substances is greater at the temperature of liquid air than at ordinary temperatures. On the whole, however, the concordance of the values of this quantity for so many metals over so wide a range of temperatures is more striking than the differences, when one considers the number of factors which might enter.

The percentage temperature coefficient required by the theory is 0·366 °/o. The experimental values for the good conductors are practically equal to this within the limits of experimental error.

---

* *Nature*, vol. LIV. p. 154; *Scientific Papers*, vol. IV. p. 232.

† *Phil. Trans.* A. vol. CCVIII. p. 381 (1908).

More Complete Theory of Conduction.

The extent to which the preceding calculations account for the observed relations between the electric and thermal conductivities of good conductors is the more surprising when one considers the approximate nature of the assumptions on which the calculations are based. For example, we assumed that all the electrons moved through equal free paths with equal velocities. Now the results of the kinetic theory considerations at the beginning of this chapter lead us to expect that instead of possessing equal velocities the electrons would have a velocity distribution in accordance with Maxwell's law. This requirement is not obviated by the possible fact that the electrons may be subjected to very intense forces; for the distribution of kinetic energy amongst the particles is, as we have seen, independent of the potential energy. Moreover the fact that the atoms, with which the electrons collide, are so massive that the velocity of an electron does not alter in magnitude during a collision with one of them makes no difference, for the encounters of the electrons with one another will ensure that Maxwell's law of distribution of velocity is established.

The following calculation* is much more general than the one we have considered. It assumes that when no external force acts on the conductor and it is all at a uniform temperature the velocity of the electrons is distributed according to Maxwell's law. When an electric force acts, or when the temperature varies from point to point of the material, we shall suppose the law of distribution to be slightly modified. We shall neglect the immediate effect of collisions between electrons compared with that of those between electrons and atomic or sub-atomic structures, and shall treat the latter as though they were immovable centres of force. The theory is therefore incomplete, since we assume that the normal distribution of velocity is that of Maxwell and at the same time that the collisions take place with particles which are immovable and are therefore unable to change the resultant velocity of the colliding particles. Thus, so far as the set of assumptions we are dealing with is concerned, a complete theory would have to take into account the occurrence of collisions between electrons, or something which produces the same effect, in order to account for the existence of Maxwell's distribution under normal conditions. The mathematical difficulties of a more complete theory, unfortunately, appear insuperable. Our method of attack will be to try to find a law of distribution of velocity, slightly different from the normal one, which will make the distribution at every point a steady one when electric or other forces act on the electrons in the metal.

---

* Cf. H. A. Lorentz, *Theory of Electrons*, p. 266; N. Bohr, *Studier over metallernes Elektrontheori*, Copenhagen (1911); O. W. Richardson, *Phil. Mag.* vol. xxiii. p. 594 (1912).

Consider the causes which tend to change the number of electrons, at any point, which have assigned components u, v, w of velocity. These are two in number, viz. (1) the free motion of the electrons from one part of the conductor to another, and (2) collisions. Let us imagine the distribution of velocity among the electrons *at any point* to be represented by a three-dimensional velocity diagram. The diagram is drawn so that the resultant velocity of each electron at the point is represented by a radius from the origin. The density of the points which are the ends of such lines and which lie within any element of volume $d\sigma = du dv dw$ of this diagram will represent the number of electrons at the given point which have velocity components between u and $u + du$, v and $v + dv$, and w and $w + dw$. In this discussion we mean, of course, by the expression the number "at any point" x, y, z the number in an infinitesimal element of volume $d\tau = dx dy dz$ which contains the point x, y, z. Let us denote the number of electrons which have velocity components u, v, w at the point x, y, z at the instant t by $f(u, v, w, x, y, z, t)\, d\sigma d\tau$. If the electric intensity in the metal is X it will give rise to an acceleration of each electron equal to Xe/m. X is supposed to be parallel to the x axis. If there were no collisions these electrons would be found in a different element $d\sigma' d\tau'$ of the velocity and space diagram at a later instant $t + dt$. Owing to the motion the new velocity coordinates would evidently be $u + X \dfrac{e}{m} dt$, v, w, and the space coordinates $x + udt$, $y + vdt$ and $z + wdt$ instead of u, v, w, x, y, z respectively. Corresponding points would be displaced equally in each diagram so that $d\sigma' = d\sigma$ and $d\tau' = d\tau$. If there were no collisions we should have as the condition for the existence of a distribution of velocity invariable with the time, that:

$$f\left(u + X\frac{e}{m}dt,\ v,\ w,\ x + u\,dt,\ y + v\,dt,\ z + w\,dt,\ t + dt\right)$$
$$= f(u, v, w, x, y, z, t).$$

The occurrence of collisions, however, makes it necessary to modify this equation. The number of electrons in the elements $d\sigma' d\tau'$ at $t + dt$ is equal, not to the number in the element $d\sigma\, d\tau$ at time t, but to the latter minus those which have disappeared from the original group plus those which have come into the group from other groups, owing to the occurrence of collisions. Since these quantities are each clearly proportional to the range of velocity and space covered by the group and to the interval of time dt, we may write them in the form $a\, d\sigma\, d\tau\, dt$ and $b\, d\sigma\, d\tau\, dt$ respectively. The condition for the existence of a steady distribution of velocity may therefore be written:

$$f\left(u + X\frac{e}{m}dt,\ v,\ w,\ x + u\,dt,\ y + v\,dt,\ z + w\,dt,\ t + dt\right)$$
$$= f(u, v, w, x, y, z, t) + (b - a)\,dt,$$

and on expanding

$$f\left(u + X\frac{e}{m}dt, \text{ etc.}\right)$$

by Taylor's theorem, we have

$$X\frac{e}{m}\frac{\partial f}{\partial u} + u\frac{\partial f}{\partial x} + v\frac{\partial f}{\partial y} + w\frac{\partial f}{\partial z} + \frac{\partial f}{\partial t} = b - a \text{ ....(31).}$$

If we can evaluate b and a in terms of f we shall evidently obtain an equation from which, if we are able to solve it, the function f may be deduced.

The evaluation of a and b is a particular case of a more general calculation given by Maxwell*. In the present instance we are treating the centres of force with which the electrons collide as immovable. This will correspond very closely with the facts on account of the small mass of the electrons. The relative velocity of the colliding particles is thus equal to the actual velocity $V = \sqrt{u^2 + v^2 + w^2}$ of the moving electron. Strictly speaking, any moving electron will be influenced by all the centres of force at every instant; but we shall suppose that when the important deflections take place the influence of one of the centres is very great compared with that of all the others. As the term collision is often understood to imply the occurrence of geometrical contact, like that between hard elastic spheres, we shall replace it by the more general term encounter, throughout the rest of this discussion. The sequence of changes which characterize an encounter is then to be represented mathematically as follows. As the previous history of the electron cannot affect the results of the calculation we can suppose it to have been moving for an indefinite time with the uniform velocity V in a straight line. As the encounter begins to occur the linear path becomes curved, but the magnitude of V is unchanged. As we are neglecting all encounters in which more than one centre of force plays an important part, the orbit of the electron will lie entirely in one plane, that which contains the direction of its original motion and the perpendicular from the centre of force upon it. If the orbit is an open one it will end by becoming asymptotic to a straight line inclined at an angle, which we shall call 2θ, to the direction of the original straight path. θ is evidently the deflection up to the apse, about which the orbit is symmetrical. If the perpendicular distance from the deflecting centre to the original straight path is b, the number of collisions made in unit time by a single particle whose speed is V, such that the distance b lies between b and $b+db$ and the plane of the orbit lies in an azimuth between ψ and $\psi+d\psi$, where ψ is measured from a fixed plane passing through the direction of V, which is taken as the axis about which ψ is measured, is $nrb\,db\,d\psi$, where n is the number of the deflecting centres in unit volume of the substance. This follows since the expression above is equal to the number of centres in the region between two circular cylinders of radii b and $b+db$, whose height is equal to the velocity V of the moving particle, which is cut off by the two planes ψ and $\psi+d\psi$ which pass through V and determine the plane of the orbit. Since there are $f(u, v, w)\,d\sigma$ particles with velocity components about u, v, w in unit volume, it follows that the number of particles which leave the group u, v, w, x, y, z in unit time owing to deflections through angles which lie between 2θ and $2(\theta+d\theta)$ in an azimuth between ψ and $\psi+d\psi$ is

---

* *Scientific Papers*, vol. II. p. 36.

$$nVf(u, v, w, x, y, z, t)\,b\,db\,d\psi\,d\sigma\,d\tau\,dt \,\ldots\ldots\ldots(32).$$

The relation between b and θ may be found from the theory of central orbits*. The law of force being K/d^s, where d is distance and K is positive when the forces are repulsive, let

$$\alpha = b\left(\frac{M_1 M_2 V^2}{K(M_1+M_2)}\right)^{\frac{1}{s-1}} \quad \text{.................. (33)},$$

where M_1 is the mass of an electron and M_2 that of one of the centres of force. Then

$$\frac{\pi}{2} - \theta = \int_0^{x'} \left\{1 - x^2 - \frac{2}{s-1}\left(\frac{x}{\alpha}\right)^{s-1}\right\}^{-\frac{1}{2}} dx \quad \text{........ (34)},$$

where x' is the least positive root of

$$1 - x^2 - \frac{2}{s-1}\left(\frac{x}{\alpha}\right)^{s-1} = 0 \quad \text{................. (35)}.$$

Evidently α is a function of θ and s only.

Since M_1/M_2 is very small we can put $(M_1+M_2)/M_1M_2 = m^{-1}$, where m is the mass of an electron. Hence

$$b\,db = \left(\frac{K}{mV^2}\right)^{\frac{2}{s-1}} \alpha\, d\alpha.$$

Thus the expression (32) may be written

$$n\left(\frac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} f(u, v, w)\, \alpha\, d\alpha\, d\psi\, d\sigma\, d\tau\, dt \quad \text{..... (36)}.$$

Now consider the reverse collisions which bring new electrons into the group u, v, w. Let u', v', w' denote the velocity components, before the encounter, of an electron which, as the result of the encounter, acquires the velocity components u, v, w, i.e. joins the specified group. Consider first the reverse encounters for which θ and ψ lie within the assigned limits considered above. Since the value of the resultant velocity is unchanged by a collision, we have

$$V' = \sqrt{u'^2 + v'^2 + w'^2} = \sqrt{u^2 + v^2 + w^2} = V.$$

The individual components u', v', w' will be different from u, v, w on account of the rotation of V through the angle 2θ. They may be written down by making use of the fact that the component of velocity along the apsidal distance has been reversed whilst the perpendicular component is unaltered. Thus if $\psi = 0$ is the plane containing V and u,

---

* Maxwell, *Scientific Papers*, vol. II. p. 36; Routh's *Particle Dynamics*, chap. VI. p. 198 (Cambridge, 1898).

$$u' = u - 2u \sin^2 \theta + \sqrt{v^2 + w^2} \sin 2\theta \cos \psi,$$

with similar expressions for v' and w'. The volume $du'dv'dw'$ of the three-dimensional velocity diagram which is occupied by the deflected points will be equal to $dudvdw$, since the new points may be obtained by reflecting the undeflected points in a plane perpendicular to the orbit and tangential to it at the apse. Thus, in considering the reverse collisions which bring extraneous electrons into the group $f(u, v, w)\, dudvdw$, the only change we require to make in (36) is the replacement of $f(u, v, w)$ by $f(u', v', w')$. Hence

$$b - a = n \left(\frac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} \iint \{f(u', v', w') - f(u, v, w)\}\, \alpha d\alpha d\psi.$$

We shall confine our discussion to the one-dimensional case in which the state of the material depends only on the x coordinate, i.e. in the electrical problem we shall suppose the electric intensity to be parallel to the axis of x and in the thermal problem we shall suppose the temperature gradient to lie in this direction. Under these restrictions equation (31), when applied to the steady state which does not change with the time, becomes

$$n \left(\frac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} \iint \{f(u', v', w') - f(u, v, w)\}\, \alpha d\alpha d\psi$$
$$= X \frac{e}{m} \frac{\partial f}{\partial u} + u \frac{\partial f}{\partial x} \quad \ldots(37).$$

The left-hand side is the rate of change of f owing to collisions, and the right-hand side that which arises from the displacement of the group of electrons as a whole.

Now when the material is at a uniform temperature T and there is no applied electromotive force, the distribution of velocity among the electrons is in accordance with Maxwell's law. Thus, if $f_0(u, v, w)$ denotes the value of the function f under these conditions, we may write

$$f_0(u, v, w) = Ae^{-hV^2} \quad \ldots\ldots\ldots\ldots(38),$$

where A and h are constant throughout the material. It is likely that this distribution will be only slightly changed by the changed conditions we are contemplating, since the new forces which come into play are always small compared with the intermolecular forces. Let us see if we can satisfy (37) by a solution of the form

$$f(u, v, w) = Ae^{-hV^2} + u\chi(V) \ldots\ldots\ldots\ldots (38\,a),$$

where the term $u\chi(V)$ is small compared with Ae^{-hV^2} and $\chi(V)$ is a function of V, the resultant velocity, only. There are two conditions which any solution will have to satisfy. These are that the total number of electrons in unit volume of the material, and their mean kinetic energy at any part which is at an assigned temperature, should both have the same values as in the uniform condition to which (38) applies. If we carry out the integrations over the whole of the velocity diagram, we find that the former is $N = A(\pi/h)^{\frac{3}{2}}$ and the latter $\frac{1}{2}m\overline{V}^2 = 3m/4h$, whether we use (38) or (38 a). Hence (38 a) satisfies the required conditions so far.

In order to determine $\chi(V)$ we make use of equation (37). Since the differences of f do not occur on the right-hand side of this equation, we shall neglect the effect of the small term $u\chi(V)$ on the right-hand side of (37). Substituting the assumed solution, the left-hand side becomes

$$n\left(\frac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} \chi(V) \int_0^\infty \int_0^\pi (u' - u)\, \alpha\, d\alpha\, d\psi$$

$$= n\left(\frac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} \chi(V) \int_0^\infty - 4\pi u \sin^2\theta\, \alpha\, d\alpha,$$

whilst the right-hand side is

$$\left(-2hAX\frac{e}{m} + \frac{\partial A}{\partial x} - V^2 A \frac{\partial h}{\partial x}\right) u e^{-hV^2}.$$

Thus

$$\chi(V) = \frac{\left(2hAX\dfrac{e}{m} - \dfrac{\partial A}{\partial x} + V^2 A\dfrac{\partial h}{\partial x}\right) e^{-hV^2}}{4\pi n \left(\dfrac{K}{m}\right)^{\frac{2}{s-1}} V^{1-\frac{4}{s-1}} \displaystyle\int_0^\infty \sin^2\theta\, \alpha\, d\alpha} \ldots\ldots\ldots\ldots (39)$$

$$= \phi V^{\frac{4}{s-1}-1}\left(2hAX\frac{e}{m} - \frac{\partial A}{\partial x} + V^2 A\frac{\partial h}{\partial x}\right) e^{-hV^2} \ldots\ldots (40),$$

where

$$\phi^{-1} = 4\pi n \left(\frac{K}{m}\right)^{\frac{2}{s-1}} \int_0^\infty \sin^2\theta\,\alpha\,d\alpha \quad \text{..............(41).}$$

The definite integral in (41) is a function of s only, and may be evaluated graphically when s is known. Thus $\chi(V)$ does not involve α nor ψ, and only involves u, v, w in the combination V. It is therefore possible to satisfy (37) by a solution of the form of (38 a), where $\chi(V)$ is a function of V only and is given by (39).

The electric current density, J, will be obtained if we multiply the number of electrons in unit volume within a specified infinitesimal range of velocities, by their component of velocity parallel to the electric intensity and by the charge each carries, and then integrate the product over all the electrons in unit volume. Thus

$$\begin{aligned} J &= 4\pi e \int_0^\infty u^2 \chi(V)\, V^2\, dV = \frac{4\pi}{3} e \int_0^\infty V^4 \chi(V)\, dV \\ &= \frac{2\pi}{3}\phi e \frac{\Gamma\left(\frac{2}{s-1}+2\right)}{h^{\frac{2}{s-1}+2}} \left[2hAX\frac{e}{m} - \frac{\partial A}{\partial x} + \left(\frac{2}{s-1}+2\right) A \frac{1}{h}\frac{\partial h}{\partial x}\right] \\ &\qquad\qquad \text{............(42),} \end{aligned}$$

remembering that

$$\int_0^\infty e^{-x} x^p\, dx = \Gamma(p+1) = p\Gamma(p).$$

The specific electrical conductivity, σ, is the coefficient of X in (42) when there is no temperature gradient in the material, i.e. when $\partial A/\partial x$ and $\partial h/\partial x$ are zero. Thus

$$\sigma = \frac{4\pi}{3}\phi\frac{e^2}{m}\frac{\Gamma\left(\frac{2}{s-1}+2\right)}{h^{\frac{2}{s-1}+1}} A \quad \text{..............(43).}$$

The thermal current density W is obtained by integrating the product of the number of electrons in unit volume, which have velocity components within a given range, by their thermal energy $\frac{1}{2}mV^2$ and by their velocity component u parallel to the temperature gradient, over all the values of u, v, w which occur.

Thus

$$W = 2\pi m \int_0^\infty u^2 \chi(V)\, V^4 dV = \frac{2\pi}{3} m \int_0^\infty V^6 \chi(V)\, dV$$

$$= \frac{\pi}{3} \phi m \frac{\Gamma\left(\frac{2}{s-1}+3\right)}{h^{\frac{2}{s-1}+3}} \left[2hAX\frac{e}{m} - \frac{\partial A}{\partial x} + \left(\frac{2}{s-1}+3\right) A \frac{1}{h}\frac{\partial h}{\partial x}\right] \qquad \text{............(43}\,a\text{)}.$$

When the electric current density J vanishes, we have from (42)

$$2hAX\frac{e}{m} - \frac{dA}{dx} = -\left(\frac{2}{s-1}+2\right) A \frac{1}{h}\frac{\partial h}{\partial x}.$$

Since, from p. 405, $h = \frac{k_0 m}{2} = \frac{m}{2RT}$ by equation (24), we see that

$$\frac{1}{h}\frac{\partial h}{\partial x} = -\frac{1}{T}\frac{\partial T}{\partial x}.$$

The thermal current, W_0, when there is no electric current, is therefore

$$W_0 = -\frac{\pi}{3}\frac{\phi m}{T} \frac{\Gamma\left(\frac{2}{s-1}+3\right)}{h^{\frac{2}{s-1}+3}} A \frac{\partial T}{\partial x},$$

and the coefficient, k, of thermal conductivity is the coefficient of $-\frac{\partial T}{\partial x}$ in this expression or

$$k = \frac{\pi}{3}\frac{\phi m}{T} \frac{\Gamma\left(\frac{2}{s-1}+3\right)}{h^{\frac{2}{s-1}+3}} A \qquad \text{..................(44)}$$

The ratio of the thermal to the electrical conductivity is

$$\frac{k}{\sigma} = \frac{1}{4}\frac{m^2}{e^2 T} \frac{\frac{2}{s-1}+2}{h^2} = \frac{2s}{s-1}\frac{R^2}{e^2} T \qquad \text{............(45).}$$

The simplicity of equation (45) is very striking. It shows that the ratio of the thermal to the electrical conductivity is independent of the number and mass of the free electrons and of the number and strength of the centres of force. In addition to R, e and T it depends only on s, the index which determines the mode of variation of the force exerted by the centres with the mutual distance. We have seen (p. 411) that the results of experiments show that for the pure metals the value of k/σ is $3R^2T/e^2$; whence we conclude that $2s/(s-1)=3$ or $s=3$. It follows that the encounters of the electrons in all metals are such as would occur with immovable particles which exert central forces varying as the inverse third power of the mutual distance.

We can estimate the order of magnitude of the strength, K, of the centres from that of the electrical conductivity of the metals by substitution in equation (43). Substituting the values of A and ϕ, we have

$$K=\frac{2\sqrt{2}}{3\pi^{\frac{3}{2}}}\,\frac{N}{n}\left(\frac{\nu R}{\nu m}\right)^{\frac{1}{2}}\frac{e^2}{\sigma}\,\frac{T^{\frac{1}{2}}}{\int_0^\infty \sin^2\phi\alpha\,d\alpha}\quad\text{............(46)},$$

where ν is the number of molecules in 1 c.c. of any gas at 0° C. and 760 mms. We do not know either the number, N, of free electrons or the number, n, of centres of force in unit volume; but we shall assume that both these quantities are of the same order of magnitude as the number, p, of atoms of the metal in unit volume. Let us put $N=\gamma p$ and $n=\delta p$; then γ and δ are numbers comparable with unity. The other data required are, in general: $\nu R=3{\cdot}72\times 10^3$ ergs/° C., $\nu m=5\times 10^{-8}$ gm. per c.c., $e=1{\cdot}6\times 10^{-20}$ E.M. unit, $T=273$°C. and $\int_0^\infty \sin^2\phi\alpha\,d\alpha$ which, when $s=3$, is equal to

$$4\pi^2\int_0^{\pi/2}\frac{\psi\cos^2\psi\,d\psi}{(\pi^2-4\psi^2)^2}=\cdot 396.$$

Taking the case of silver as an illustration, $\sigma=6\times 10^{-4}$ E.M. unit and $\delta K/\gamma=8{\cdot}5\times 10^{-26}$.

This result is of additional interest because several other considerations point to the occurrence of forces varying inversely as the cube of the distance as an important feature of atomic structure. Thus J. J. Thomson* and Jeans† have shown that some of the laws which govern the emission of thermal radiation would follow if the collisions of the electrons were with centres of force obeying the inverse third power law. The strength, K, of these centres can be estimated from the constants in the radiation formulae. The mean of the estimates given by Thomson and Jeans is $K = 2{\cdot}5 \times 10^{-27}$. This agrees satisfactorily with that given by the electrical conductivity.

---

* *Phil. Mag.* vol. XIV. p. 217 (1907); vol. XX. p. 238 (1910).

† *Ibid.* vol. XX. p. 642 (1910).

If the centres of force are due to the occurrence of electrical doublets inside the atom, it is possible to have a state of steady motion in which an electron revolves about the axis of the doublet. The energy of the steady motion, as well as that of the small oscillations about it, is proportional to the frequency of the rotation. Thomson* has suggested this result as an explanation of the emission of electrons by the action of light (cf. p. 469); in which case, as we shall see in the next chapter, the kinetic energy of the emitted electrons is a linear function of the frequency of the exciting light. Another interesting property of these orbits is that the moment of the magnet to which they are equivalent depends only on the moment of the doublets about which they are rotating. Since the universality of the law connecting radiation and temperature suggests that the strength of these doublets is independent of the matter in which they occur, they would furnish an explanation of the atomic magnets, or magnetons, whose existence has been inferred by Weiss (cf. also p. 395).

Electrical Conductivity as a Function of Temperature.

Although the electron theory has been very successful in explaining the laws of Wiedemann and Franz and of Lorenz, which are obeyed by the ratio of the thermal to the electrical conductivity, it has not been so successful in accounting for the behaviour of these two properties individually. In point of fact, for the pure metals the thermal conductivity is practically independent of the temperature, whereas the electrical conductivity is approximately inversely proportional to the absolute temperature, except at low temperatures. Since

$$\sigma = \frac{2\sqrt{2}}{3\pi^{\frac{3}{2}}} \frac{N}{n} \left(\frac{R}{m}\right)^{\frac{1}{2}} \frac{e^2}{K} \frac{T^{\frac{1}{2}}}{\int_0^x \sin^2\phi\, a\, da} \text{ ....(47)},$$

even if we assume the index s which determines the law of force to be practically independent of T, the observed law σ varies as

---

[*] *Phil. Mag.* vol. xx. p. 243 (1910).

T^{-1} might arise from an appropriate variation of one or more of the quantities N, n and K with T. At present there is no means of finding the connection of K, N and n separately with T.

Recent experiments by Kammerlingh Onnes and his collaborators have shown that the resistance of the pure metals becomes exceedingly small at very low temperatures. This result cannot be said to be definitely indicated by the formulae which we have just developed. It is, however, not necessarily in conflict with them. For example we might identify the centres, with which the electrons collide, with the vibrators which take part in thermal phenomena. The most obvious physical interpretation of these vibrators is to regard them as electrical doublets. The doublets might arise, for example, by the vibration or steady motion of electrons about an equilibrium position. They would exert forces on the moving electrons which would vary inversely as the cube of the mutual distance, a result which, as we have seen, is required by several considerations. It follows from Einstein's theory of specific heats that the energy of the vibrators approaches zero exponentially as the absolute zero of temperature is approached. Now the average moment of a doublet constituted in this way is proportional to the square root of its energy: so that the moment or strength of the centres will also approach zero exponentially as the temperature is reduced. Since there is no reason for supposing that the number of free electrons approaches zero at so fast a rate as this, the conductivity would become infinite at very low temperatures.

Such a view is in satisfactory agreement with the experimental measurements. By assuming that K times the product on the right-hand side of (47) varies as $T^{-\frac{1}{2}}$, which gives the right variation of σ with T at high temperatures, and that the variation of K with T depends upon the contained heat energy, as deduced from the specific heat measurements, in the manner just indicated, I find that the calculated temperature at which the electrical resistance disappears agrees quite accurately with that given by Kammerlingh Onnes in the case of mercury and is not far from the values given for lead and gold. The way in which the calculated electrical resistance varies with the temperature is also very similar to that found experimentally at these low temperatures. The same results would not follow from Planck's later theory of specific heats mentioned in Chapter XV. According to that view the energy of the vibrators does not approach the value zero at low temperatures but the finite value $h\nu/2$. This would make K approach a finite value, so that the resistance would not vanish.

Thermoelectric Phenomena.

As is well known, if a circuit which consists of wires of two different materials is constructed, a current will flow round it without the assistance of a battery, if the two junctions are at different temperatures. There is no current when the two junctions are at the same temperature. As there is no observable change in the nature or composition of the materials constituting the circuit, the electromotive force which drives these currents must be derived from the available thermal energy. By causing the electric currents to perform mechanical work we could evidently construct a heat engine out of a circuit of this kind. As long ago as 1854 Lord Kelvin showed that valuable information about thermoelectric phenomena might be obtained by the application of the principles of thermodynamics to thermoelectric circuits.

The thermoelectromotive force of a circuit consisting of the two conductors A and B may be denoted by E_{AB}. It has been found that for a very large number of pairs of substances E_{AB} can be represented within the limits of accuracy of experimental observation by the comparatively simple empirical formula

$$E_{AB} = a(T_2 - T_1) + \tfrac{1}{2}b(T_2^2 - T_1^2),$$

where T_2 and T_1 are the temperatures of the two junctions and a and b are constants. The differential coefficient of E_{AB} with respect to the upper temperature T_2 is called the thermoelectric power of the circuit. It is clearly a linear function of the temperature within the limits of accuracy of the formula above. The thermoelectromotive forces between circuits of different metals terminating at the same pair of temperatures are connected by the relation

$$E_{AB} + E_{BC} = E_{AC} \quad\text{..................}(48).$$

This is evident since the circuit composed by the addition of AB and BC in the order of the letters only differs from AC by the inclusion of a single wire of the material B, the temperature of which varies from one extreme to the other and back again. As inequality of temperature in a closed circuit of a single uniform material does not give rise to any thermoelectromotive force, this cannot cause any difference from the effect given by AC simply.

In addition to the thermoelectromotive force caused by a difference of temperature there are the converse reversible heating and cooling effects which are produced by the flow of an electric current. We have thus to consider the Peltier effect, which is the heat liberated when an electric current flows across the junction between two different metals, and the Thomson effect, which is the heat developed reversibly when an electric current flows along an unequally heated bar. These are measured by the respective coefficients P and σ. Both of these refer to the amount of heat liberated by the passage of unit quantity of electricity. In specifying σ the electricity has to flow against one degree difference of temperature, the directions of the electric and thermal gradients being coincident. σ is positive when positive current flowing in the direction of increasing temperature causes an absorption of heat.

The application of the conservation of energy to a thermoelectric circuit gives

$$Ei = Ri^2 + i\Sigma P + i\int \sigma \, dT$$

if R is the resistance of the circuit. When i is made very small the Joulian development of heat Ri^2 vanishes in comparison with the other terms, so that the reversible quantities satisfy the equation

$$E = \Sigma P + \int \sigma \, dT \text{ ....(49).}$$

Similarly by applying the second law of thermodynamics in the form $\int \frac{dQ}{T} = 0$ to the reversible heat production, we have

$$\Sigma \frac{P}{T} + \int \frac{\sigma}{T} dT = 0 \text{ ....(50).}$$

It follows from these equations that

$$\frac{\partial E}{\partial T} = \frac{P}{T} \quad \text{and} \quad \sigma_B - \sigma_A = T\frac{\partial}{\partial T}\left(\frac{P}{T}\right) \text{ ....(51).}$$

These conclusions have been confirmed in many particulars, and there are no experimental results which can be said with certainty to conflict with them. The values of the different thermoelectric quantities may be derived from the expressions given on pp. 420—421, for the rate of transportation of electricity and heat. These expressions were derived on the hypothesis that the law of force between the electrons and the centres with which they collide varied as a power of the mutual distance. They will hold strictly only if the centres are at distances apart which are large compared with those within which the forces are appreciable. The two principal reasons for this are: (1) if this condition is not satisfied all the particles are under collision conditions at every instant and the physical foundation of the calculation disappears, and (2) as the potential energy varies very rapidly from point to point in the neighbourhood of any electron the quantity A will be subject to corresponding sharp variations (cf. equation (15), p. 403), and the actual A can therefore scarcely be regarded as differentiable.

In order to embrace these conditions we shall generalize the former calculations a little at the sacrifice of a certain amount of rigour. We shall assume that the average value of $\boldsymbol{A}$ is differentiable however complicated the internal fields of force may be and that, for the steady flow parallel to the axis of u, f is always of the form $Ae^{-hV^2} + u\chi(V)$, where χ involves u, v and w through V only. By working through a calculation similar to that already carried out, it becomes clear that, provided the medium is isotropic, $\chi(V)$ will be of the form

$$\left(2hAX - \frac{dA}{dx} + V^2 A \frac{dh}{dx}\right) \psi e^{-hV^2},$$

where the function ψ involves only the mass, m, of the electrons, the number, n, of the centres in unit volume, the law of force and V. In this way the equations for the currents of electricity and heat may be written:

$$i = \frac{4\pi}{3} e \left\{\left(2hAX - \frac{dA}{dx}\right) J_4 + A \frac{dh}{dx} J_6\right\} \quad \text{............(52)}$$

$$= \frac{\sigma_0}{e} \left\{mX - RT \frac{d \log A}{dx} + \mu R \frac{dT}{dx}\right\} \quad \text{...............(53)},$$

$$W = \frac{2\pi}{3} m \left\{ \left(2hAX - \frac{dA}{dx} \right) J_6 + A \frac{dh}{dx} J_8 \right\} \quad \text{.............(54)}$$

$$= -\frac{RT}{e^2} \sigma_0 \mu \left\{ mX - RT \frac{d \log A}{dx} + \mu' R \frac{dT}{dx} \right\} \quad \text{.......(55)}$$

$$= -\frac{RT}{e} \mu i - k \frac{dT}{dx} \quad \text{..................................(56),}$$

where

$$J_n = \int_0^\infty V^n \psi e^{-hV^2} dV,$$

$$\sigma_0 = \text{specific electrical conductivity} = \frac{4\pi}{3} \frac{e^2}{RT} A J_4,$$

$$\mu = -\tfrac{1}{2} \frac{m}{RT} J_6/J_4, \quad \mu' = -\frac{m}{2RT} J_8/J_6,$$

$$k = \text{thermal conductivity} = \frac{\pi}{3} \cdot \frac{A}{R} \frac{m^2}{T^2} \left\{ J_8 - \frac{J_6^2}{J_4} \right\},$$

and $$k/\sigma_0 = \frac{m^2}{e^2} \frac{1}{T} \frac{1}{4} \left\{ \frac{J_8}{J_4} - \left(\frac{J_6}{J_4} \right)^2 \right\} = \frac{R^2}{e^2} T\mu (\mu' - \mu).$$

If the centres are far enough apart and act on one another with forces which vary as the inverse *s*th power of the distance, we have, as before,

$$\mu = -\left(\frac{2}{s-1} + 2 \right) \quad \text{and} \quad \mu' = -\left(\frac{2}{s-1} + 3 \right).$$

In the equations above, X is the mechanical force on an electron. If X_0 is the electric intensity, then $eX_0 = mX$.

The thermoelectromotive force E round any circuit is the value of $\int X_0 dx$ which is required to reduce the current i to zero, when there are no batteries in the circuit. Thus from (53)

$$E = \int X_0 dx \quad \text{when} \quad i = 0$$

$$= \frac{R}{e} \int \left\{ T \frac{d \log A}{dx} - \mu \frac{dT}{dx} \right\} dx$$

$$= \frac{R}{e} \left\{ \left| T \log A \right| - \int (\log A + \mu) \, dT \right\} \quad \text{........(57),}$$

after integrating by parts. If the integral is taken round a closed circuit of two metals with varying temperatures, the integrated part has the same value at the identical limits. The unintegrated part, since the integrand is not necessarily a perfect differential, will not necessarily be zero. If the temperatures of the junctions are T' and T_0, it may be written

$$E_{12} = \frac{R}{e}\int_{T_0}^{T'} \left\{\log\frac{A_2}{A_1} + (\mu_2 - \mu_1)\right\} dT \quad \text{.........(58)},$$

where the suffixes 1 and 2 denote the respective materials. The thermoelectric power is

$$\frac{\partial E_{12}}{\partial T'} = \frac{R}{e}\left\{\log\frac{A_2}{A_1} + (\mu_2 - \mu_1)\right\}_{T'} \quad \text{..................(59)}.$$

The Thomson coefficient, σ, may be obtained as follows:—The heat developed per unit area and thickness dx in the direction of the current is equal to the work done by the electric force inside the volume + the stream of energy flowing in − the stream of energy flowing out. This is

$$\left(X_0 i - \frac{\partial W}{\partial x}\right) dx = \left[\frac{i^2}{\sigma_0} + i\left\{\frac{RT}{e}\frac{\partial}{\partial x}(\log A + \mu)\right\} + \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)\right] dx \quad \text{............(60)},$$

from (53) and (56). The first term on the right in (60) represents the Joule heating effect and the third term is independent of the electric current. The middle term alone reverses with the electric current and is therefore the heat production owing to the Thomson effect. The Thomson coefficient σ, which is the amount of heat absorption per unit current per unit time per unit rise of temperature, is therefore

$$\sigma = -\frac{RT}{e}\frac{\partial}{\partial T}(\log A + \mu) \quad \text{..................(61)}.$$

The value of the Peltier effect may be obtained by applying the same equation to the passage of a current across the junction between two metals at the same temperature instead of considering the flow along a bar of the same metal with varying temperature. The magnitude of the Peltier coefficient P_{12} is given by

$$eP_{12} = RT\left(\log\frac{A_1}{A_2} + \mu_1 - \mu_2\right) \quad \text{..............(62)}.$$

This is not quite equal to the work done by the electric force in the neighbourhood of the separating surface. For we see from

(53) that when there is no difference of temperature $\left(\frac{\partial T}{\partial x}=0\right)$, as i approaches zero X_0 approaches $\frac{RT}{e}\frac{\partial \log A}{\partial x}$; so that the work done by the electric force at the boundary is $RT \log \frac{A_1}{A_2}$. The difference between this and eP_{12} is due to the fact that, when the law of force is different in different metals, the amount of kinetic energy transported by the electrons which carry a given current is different. As the ratio of the thermal and electric conductivities indicates that the law of force is very nearly the same for the pure metals which are good conductors, it is probable that $\mu_1 - \mu_2$ is not large. It may, however, be of the same order of magnitude as the observed Peltier and Thomson effects.

The values given by (58), (59), (61) and (62) satisfy the equations (49)—(51) of Lord Kelvin's thermodynamic theory. A different method of deducing some of the thermoelectric formulae will be considered in the next chapter.

Conductivity for Periodic Forces.

The behaviour of the electrons in metals under alternating forces has been considered by various writers. The following investigation, due to H. A. Wilson*, is an elaboration of a method originally given by Jeans†. The electrons instantaneously present in any given volume may be divided into groups characterized by a particular velocity of agitation V. The number of collisions of the electrons with one another is regarded as small compared with the number of collisions between electrons and atoms. Owing to the relatively small mass of the electrons the latter class of collisions will have very little effect in changing the magnitude of the velocity V. Thus the electrons in any given group will remain in that group throughout a large number of collisions. Let dN denote the number of electrons in the group characterized by the value V and let u denote the average component of velocity of this group in the direction of the electric intensity X. Consider the rate at which the momentum of the group is changing. The group is gaining momentum from the applied field at a rate which is equal to the force which acts on the electrons in the group. The magnitude of this is $XedN$. At the same time the momentum thus acquired is being dissipated by collisions at a rate $mu\beta dN$, where β depends on V, on the strength of the centres and on the law of force which governs the collisions. The value of β can be found by considering the deviation of the electrons produced by collisions*. Thus in general the change of momentum of the group with time will be in accordance with the differential equation:

---

* *Phil. Mag.* VI. vol. xx. p. 835 (1910).

† *Ibid.* vol. xvii. p. 773 (1909).

$$\frac{d}{dt}(mudN) = XedN - um\beta dN \quad \text{............(63).}$$

When we are dealing with direct currents a steady motion is soon established which is independent of the time. The value of u for this case is thus obtained by putting the left-hand side of equation (63) equal to zero. The current density is $e\int udN = \sigma X$, where σ is the specific electrical conductivity. When the same assumptions are made about the nature of the collisions and about the law of distribution of velocity among the electrons (dN as a function of V), this leads to values of σ in agreement with those given by other methods for direct currents.

To find the conductivity for periodic forces let $X = X_0 \cos pt$. After dividing by dN, (63) becomes

$$m\frac{du}{dt} + m\beta u = X_0 e \cos pt.$$

As in the theory of dispersion, the solution of this equation which corresponds to stationary conditions is the particular integral

$$u = X_0 \frac{e}{m} \frac{\cos(pt - \delta)}{\sqrt{p^2 + \beta^2}},$$

where $\tan\delta = p/\beta$. The mean value of u for all the electrons is

$$\bar{u} = \frac{1}{N}\int udN.$$

Since all the work done by the electromotive force is converted into heat the rate of heat production is the mean value of

$$Ne\bar{u}X = Ne\bar{u}X_0 \cos pt.$$

It is also equal to $\frac{1}{2}\sigma_p X_0^2$, where σ_p is the conductivity for forces whose frequency is p. This statement may be taken as the definition of σ_p. The mean value is to be taken over a time which is large compared with $2\pi/p$. Thus $\frac{1}{2}\sigma_p X_0^2$ is equal to the mean value of

---

* Cf. H. A. Wilson, *loc. cit.*

$$\frac{e^2}{m} X_0^2 \int \frac{\cos pt \cos (pt - \delta)}{\sqrt{p^2 + \beta^2}} dN \quad \text{and} \quad \sigma_p = \frac{e^2}{m} \int \frac{\beta dN}{p^2 + \beta^2} \ldots (64).$$

In this expression N and β are functions of V and the integral is one with respect to V between the limits 0 and $+\infty$. As an illustration we may consider the simplest possible supposition that we can make, namely that the electrons all have equal velocities. The integral then reduces to a sum over the N electrons for all of which β has the same value. For zero frequency ($p = 0$) the value of σ_p then becomes

$$\sigma_0 = \frac{e^2}{m} \frac{N}{\beta},$$

and

$$\sigma_p = \frac{e^2}{m} \frac{\beta N}{p^2 + \beta^2} = \sigma_0/(1 + p^2/\beta^2) = \sigma_0 \Big/ \left(1 + p^2 \frac{m^2 \sigma_0^2}{N^2 e^4}\right) \ldots (65).$$

All the quantities except N on the right-hand side of this equation are known with considerable accuracy, and σ_p can be deduced from experiments on the optical properties of metals. Schuster* has applied equation (65) to the optical data accumulated by Drude†. In this way he finds that for all the commoner metals the number of free electrons in a given volume is from one to three times as great as the number of atoms present. It may be that these estimates are subject to errors arising from the occurrence of selective optical absorption, a phenomenon which is disregarded in the theory above; but it does not seem likely that the number can be much smaller than this. For the experiments of Rubens and Hagen with infra-red radiation show that σ_p does not differ appreciably from σ_0 until the visible spectrum is approached rather closely. This would not be true if N had values much smaller than those calculated by Schuster.

Thermal Radiation.

The theory of the motion of electrons in metals is of importance from another point of view because it helps us to form a judgment as to the mechanism of the emission of electromagnetic thermal radiation from hot bodies. We saw in Chap. XV that the spectral distribution of energy in the radiation inside an enclosure in equilibrium with matter was independent of the nature of the matter and determined solely by the temperature of the enclosure. It follows that if we can imagine a type of matter which is sufficiently real and at the same time sufficiently simple in its behaviour to admit of our calculating the properties of the radiation which is in equilibrium with it at a given temperature, we shall have solved the radiation problem. Such a possibility would seem to be offered by the theory of the motion of electrons in metals. We know that during the collisions the electrons are accelerated and that when electrically charged particles are accelerated they emit radiation. The accelerations of the other constituents of the system are probably negligible in comparison with those of the electrons; so that it would seem that we are not likely to be led into serious error if we attribute all the radiation to the motion of the electrons. A very general knowledge of the nature of the motions enables us to analyse the radiation thus emitted into its constituent frequencies by means of Fourier's series. In this way we can arrive at a knowledge of the way in which energy of assigned frequency is being emitted by the moving electrons. In the steady state an equal amount of radiant energy will be absorbed by the system. The absorption of energy occurs through the Joule heating effect of the electric currents established by the electric intensity in the electromagnetic waves. If the intensity in the waves of frequency p is X_p the absorption of energy per unit volume per unit time is the mean value of $\sigma_p X_p^2$, where σ_p is the conductivity for currents of frequency p. As we have seen σ_p is a function of p. By equating the amount of emitted energy of a given frequency to that absorbed we arrive at an expression for the steady energy density of given frequency.

---

* *Phil. Mag.* VI. vol. VII. p. 151 (1904).

† *Ann. der Physik*, vol. XXXIX. p. 537 (1890); vol. XLII. p. 189 (1891).

This method of calculating the distribution of energy in the black body spectrum was first used by Lorentz[*]. His calculations were confined to waves of low frequencies. The method has since been extended by various writers, including J. J. Thomson[†],

---

[*] Cf. *Theory of Electrons*, p. 80; *Amsterdam Proc.* 1902-3, p. 666.

[†] *Phil. Mag.* VI. vol. XIV. p. 217 (1907); vol. XX. p. 238 (1910).

Jeans*, H. A. Wilson† and Bohr‡. It appears that if the same assumptions are consistently carried through the calculations the distribution of energy found is that given by the law of Jeans and Lord Rayleigh. As we have seen, this law is not in accordance with the experimental facts; so that the method is not adequate to the solution of the problem. In fact, Jeans§ and McLaren‖ have shown pretty conclusively that any dynamical method leads inevitably to Jeans's law, so that a dynamical foundation of the theory of radiation does not seem to be possible.

Galvanomagnetic and Thermomagnetic Phenomena.

A number of interesting phenomena are observed when a conductor carrying an electric or thermal current is placed in a magnetic field. The effects are conveniently classified according to whether they are exerted across or along the primary current. The transverse effects are as follows:

(1) Hall Effect. When an electric current flows across the lines of force of a magnetic field an electromotive force is observed which is at right angles to both the primary current and the magnetic field.

(2) von Ettingshausen's Effect. Under the like circumstances a temperature gradient is observed which has the opposite direction to the Hall electromotive force.

(3) Nernst and von Ettingshausen's Effect. When heat flows across the lines of magnetic force there is an electromotive force in the mutually perpendicular direction.

(4) Leduc and Righi's Effect. Under the like conditions there is a transverse temperature gradient.

The transverse effects are all proportional to the vector product of the intensity H of the magnetic field (at any rate for sufficiently small fields) and the primary current i of heat or electricity. Thus any of the effects of amount E may be considered to be measured by the coefficients $\pi_1, \pi_2, \pi_3, \pi_4$, where, for each suffix,

$$E = \pi H i \qquad \text{(66)}.$$

---

* *Phil. Mag.* vol. XVII. p. 774 (1909); vol. XVIII. p. 209 (1909).

† *Ibid.* vol. XX. p. 835 (1910).

‡ *Studier over metallernes Elektrontheori*, Copenhagen (1911).

§ *Phil. Mag.* VI. vol. XVII. p. 229 (1909).

‖ *Ibid.* vol. XXI. p. 15 (1911).

The suffixes correspond to the numbers which precede the effects as enumerated above. π is taken to be positive if, when the primary current flows in the positive direction along the x axis and H is in the positive direction along the y axis, the resulting electric or temperature gradient is in the positive direction along the z axis, the arrangement of the axes being such that a right-handed screw travelling along Ox would rotate from Oy to Oz. This choice of signs is not, however, the one which occurs in the literature of the subject. A series of measurements by Zahn of the various coefficients for different conductors led to the values in the following table:

Conductor	π_1	π_2	π_3	π_4
Iridium	$+4{\cdot}02\times10^{-4}$	$-5{\cdot}5\times10^{-8}$	-5×10^{-6}	too small to measure
Palladium I	$-6{\cdot}91\times10^{-4}$	$+3{\cdot}3\times10^{-8}$	$+1{\cdot}27\times10^{-4}$	
Palladium II	$-11{\cdot}12\times10^{-4}$	$+1{\cdot}8\times10^{-8}$	$+0{\cdot}51\times10^{-4}$	
Platinum	$-1{\cdot}27\times10^{-4}$	$+2{\cdot}1\times10^{-8}$	very small	
Copper	$-4{\cdot}28\times10^{-4}$	$+23{\cdot}2\times10^{-8}$	$-2{\cdot}7\times10^{-4}$	
Silver	$-8{\cdot}97\times10^{-4}$	$+40{\cdot}4\times10^{-8}$	$-4{\cdot}3\times10^{-4}$	
Zinc	$+10{\cdot}4\times10^{-4}$	$-12{\cdot}9\times10^{-8}$	$-2{\cdot}4\times10^{-4}$	
Iron	$+10{\cdot}8\times10^{-4}$	-39×10^{-8}	$-10{\cdot}5\times10^{-4}$	$+5{\cdot}7\times10^{-8}$
Steel	$+133{\cdot}6\times10^{-4}$	$-68{\cdot}7\times10^{-8}$	$-16{\cdot}6\times10^{-4}$	$+6{\cdot}7\times10^{-8}$
Nickel I	$-46{\cdot}9\times10^{-4}$	$+20\times10^{-8}$	$+13\times10^{-4}$	$-2{\cdot}8\times10^{-8}$
Nickel II	-125×10^{-4}	$+55\times10^{-8}$	$+35{\cdot}5\times10^{-4}$	$-17{\cdot}6\times10^{-8}$
Antimony	$+2190\times10^{-4}$	-202×10^{-8}	$+176\times10^{-4}$	-134×10^{-8}

It is clear from these numbers that the phenomena are very complicated. The sign and magnitude of the observed effects exhibit no obvious relation to any known property of the materials concerned. In the case of bismuth, a conducting material which can be obtained in the form of large crystals, it has been found that the Hall coefficient changes sign as the orientations of the primary current and the magnetic field are changed with reference to the crystal axes.

The longitudinal effects are different according as the magnetic field is across or along the direction of the primary current. With a transverse magnetic field the following longitudinal effects have been observed. Each corresponds to the transverse effect with the same number prefixed.

(1) When a conductor is placed in a transverse magnetic field a change in its specific resistance is found to occur.

(2) There is a Peltier effect at the junction between transversely magnetized and unmagnetized material.

(3) There is a thermoelectromotive force between transversely magnetized and unmagnetized material.

(4) There is an alteration in the thermal conductivity in a transverse magnetic field.

All these effects also occur in a longitudinal magnetic field, but as a rule they are then smaller. In both cases (2), (3) and (4) have been observed only with bismuth and the ferromagnetic metals. The change of resistance in a magnetic field, also, is much larger with these metals than with any of the others.

Since the change of resistance cannot depend on the sign of H, provided the material is isotropic, it follows that for small fields the change δR in the resistance R must satisfy the equation $\delta R/R = AH^2$, where A is a constant. This equation is satisfied over a considerable range of values of H by the metals which exhibit only small changes. The following values of A have been determined by Patterson:

Cadmium	$2{\cdot}82 \times 10^{-12}$	Copper	$0{\cdot}26 \times 10^{-12}$
Zinc	$0{\cdot}87 \times 10^{-12}$	Tin...............	$0{\cdot}23 \times 10^{-12}$
Gold	$0{\cdot}37 \times 10^{-12}$	Palladium	$0{\cdot}11 \times 10^{-12}$
Silver	$0{\cdot}26 \times 10^{-12}$	Platinum	$0{\cdot}06 \times 10^{-12}$

The effects are much larger at low temperatures. Thus Laws found the following values of $10^{12} \times A$:

Temperature	−186° C.	+18° C.	+50° C.	+100° C.
Cadmium	51	2·60	1·70	0·98
Zinc	18	0·88	0·58	—

With crystals the effect depends on the direction of the current with reference to the crystal axes, and with the ferromagnetic substances the intensity of magnetization appears to be at least as important as the magnetic intensity.

Theory of Galvanomagnetic Effects.

The electron theory cannot be said to have been very successful, as yet, in unravelling these complicated magnetic phenomena. No doubt one of the factors having an influence is the deflexion of the paths of the moving electrons which is produced by an external magnetic field. The magnitude of the Hall effect which would be expected to arise in this way may be calculated as follows:

We shall suppose the Hall effect to be measured by the external electromotive force which it is necessary to introduce in order to reduce the deflexion of a galvanometer again to zero. The galvanometer is connected across two points at unit distance apart which were at the same potential before the magnetic field was applied. Let the primary current i be along the axis of x, the Hall current j along the axis of y and the magnetic intensity H along the axis of z. The components of the electric intensity at any instant are X, Y, 0. If x, y, z are the coordinates of an electron, its equations of motion are

$$m\ddot{x} = eX - \frac{H}{c}e\dot{y}, \qquad m\ddot{y} = eY + \frac{H}{c}e\dot{x}, \qquad m\ddot{z} = 0 \quad \ldots\ldots(67).$$

Thus

$$m\dot{x} = eXt - \frac{H}{c}ey + mu \quad \ldots\ldots\ldots\ldots\ldots\ldots(68),$$

$$m\dot{y} = eYt + \frac{H}{c}ex + mv \quad \ldots\ldots\ldots\ldots\ldots\ldots(69),$$

if we take the origin of time to be the instant at which the last collision occurred, and put $x = y = 0$, $\dot{x} = u$ and $\dot{y} = v$ when $t = 0$. From (68)

$$x = \frac{1}{2}\frac{e}{m}Xt^2 - \frac{H}{mc}eyt + ut,$$

and from (69)

$$\dot{y} = \frac{e}{m}Yt + \frac{H}{c}\frac{e}{m}\left\{\frac{1}{2}\frac{e}{m}Xt^2 - \frac{H}{c}\frac{e}{m}yt + ut\right\} + v \quad \ldots(70).$$

If τ is the time between two collisions, assumed for simplicity to be the same in all cases, the average value of $\dot{y}$ is

$$\bar{\dot{y}} = \frac{1}{\tau}\int_0^\tau \dot{y}\,dt = \frac{e}{m}Y\frac{\tau}{2} + \frac{H}{c}\frac{e}{m}\left\{\frac{1}{6}\frac{e}{m}X\tau^2 + \frac{1}{2}\overline{\left(u - \frac{H}{c}\frac{e}{m}y\right)}\tau\right\} + \bar{v}.$$

Now $\bar{u} = \bar{v} = 0$, since these quantities are as likely to be negative as positive, and $\frac{H}{c}\bar{y}$ may be neglected compared with $\frac{1}{3}X\tau$, so that

$$\bar{\dot{y}} = \frac{1}{2}\frac{e}{m}\tau\left\{Y + \frac{1}{3}\frac{H}{c}\frac{e}{m}X\tau\right\}.$$

The Hall current density j is $= ne\dot{y}$, so that, for this to vanish,

$$Y = -\frac{1}{3}\frac{H}{c}\frac{e}{m}X\tau.$$

Also
$$\bar{\dot{x}} = \frac{1}{2}\frac{e}{m}X\tau,$$

if we neglect $\bar{y}$ as before.

Therefore
$$i = ne\bar{\dot{x}} = \frac{1}{2}n\frac{e^2}{m}X\tau,$$

or
$$Y = -\frac{2}{3}\frac{1}{c}\frac{H}{ne}i.$$

The Hall coefficient π_1 is

$$-\frac{Y}{Hi} = +\frac{2}{3}\frac{1}{ne} \quad \text{.....................(71)},$$

if all the quantities are in electromagnetic units.

Several methods of estimating n lead to values of the order 10^{22}, and putting $e = -10^{-20}$ E.M.U. approx., we see that the order of magnitude of π_1 is -10^{-2}. To the values which are given on p. 435 the following, which have been found by different observers, may be added for comparison:

Metal ...	Bi	Sn	Pb	Sb	Te	Na	Cu
π_1	-9.0	-0.00004	$+0.00009$	$+0.192$	$+530$	-0.0025	-0.0005

According to the theory above, the Hall coefficient should be negative for all substances. The positive coefficients exhibited by Lead, Antimony, Tellurium, Iridium, Zinc and Iron might be taken to indicate the presence of free positive electrons if there were any independent evidence of their occurrence. A theory along these lines has been worked out by Drude*, but it cannot be considered to have been very successful. It seems more likely that the effects are all due to negative electrons, but that the theory only takes account of part of the phenomena. It seems probable that the magnetic field affects the conditions of equilibrium of the electrons and the way in which they are emitted by the parent atoms. Suggestions towards a theory of this kind have been made by Sir J. J. Thomson†. It is also probable that the current-carrying electrons are deflected by the magnetic fields of the electrons revolving within the atoms and that the orientation of the orbits of revolution is affected by the external magnetic field.

---

* *Ann. der Physik*, vol. III. p. 369 (1900).

† *Corpuscular Theory of Matter*, chap. v. New York (1907).

In the case of cuprous iodide, a compound substance which exhibits fairly typical electronic conductivity, it is possible, by adding varying amounts of iodine in excess of that which is normally present, to cause the conductivity to vary over a wide range. Presumably this alters the concentration of the free electrons without changing other important factors appreciably; for Steinberg[*] has shown that the Hall coefficient for this substance is directly proportional to the specific resistance. It is therefore inversely as n in accordance with equation (71).

The change of resistance in a transverse magnetic field may be calculated on the electron theory in the following manner:

Using the same notation as in the theory of the Hall effect and neglecting the term Hey/c, because it is small when averaged, we have

$$\ddot{x} = \frac{e}{m}X - \frac{H}{c}\frac{e}{m}\left\{\frac{H}{c}\frac{e}{m}\left(\frac{1}{2}\frac{e}{m}Xt^2 + ut\right) + v\right\},$$

$$\dot{x} = \frac{e}{m}Xt - \frac{H}{c}\frac{e}{m}\left\{\frac{H}{c}\frac{e}{m}\left(\frac{1}{6}\frac{e}{m}Xt^3 + \frac{1}{2}ut^2\right) + vt\right\},$$

$$\bar{\dot{x}} = \frac{1}{\tau}\int_0^\tau \dot{x} = \frac{1}{2}\frac{e}{m}X\tau - \frac{H}{c}\frac{e}{m}\left[\frac{H}{c}\frac{e}{m}\left(\frac{1}{24}\frac{e}{m}X\tau^3\right)\right],$$

$$i = ne\bar{\dot{x}} = \frac{n}{2}\frac{e^2}{m}\left\{\tau - \frac{H^2}{12c^2}\frac{e^2}{m^2}\tau^3\right\}X.$$

If there is no magnetic field $H = 0$ and

$$i_0 = \frac{n}{2}\frac{e^2}{m}\tau_0 X = \sigma X.$$

It is important to notice that in the presence of a magnetic field the free time τ will not be the same as τ_0, because the magnetic field affects the curvature and therefore the average length of the free paths, and it may also affect the orientation of the systems with which the electrons collide[†]. Calling $\delta\tau$ the change in τ thus produced, the increment δi in the current, effected by the magnetic field H, is

---

[*] Cf. Baedeker, *Elektrischen Erscheinungen, etc.* pp. 104, 123.

[†] Cf. E. P. Adams, *Phys. Rev.* vol. xxiv. p. 428 (1907).

$$\delta i = \frac{n}{2}\frac{e^2}{m}\left\{\delta\tau - \frac{1}{12}\frac{H^2}{c^2}\frac{e^2}{m^2}\tau^3\right\} X = \delta\sigma\, X.$$

If $\rho = 1/\sigma$ is the specific resistance

$$\frac{\delta\rho}{\rho} = -\frac{\delta\sigma}{\sigma} = -\frac{1}{\tau_0}\left\{\delta\tau - \frac{1}{12}\frac{H^2}{c^2}\frac{e^2}{m^2}\tau^3\right\}.$$

Since the difference between τ and τ_0 is small compared with either of them, this may be written

$$\frac{\delta\rho}{\rho} = -\frac{\delta\sigma}{\sigma} = -\frac{\delta\tau}{\tau_0} + \frac{1}{3}\frac{H^2}{c^2}\left(\frac{\sigma}{ne}\right)^2 \quad \ldots\ldots\ldots\ldots(72).$$

If we may neglect $\delta\tau/\tau_0$ this equation may be used to determine n, since all the quantities are known. Using this method, which is due to Sir J. J. Thomson, Patterson* found that his measurements of the change of resistance led to the following values of n:

Metal ...	Pt	Au	Sn	Ag	Cu
	$1{\cdot}4\times10^{22}$,	$2{\cdot}2\times10^{22}$,	$4{\cdot}5\times10^{21}$,	$3{\cdot}6\times10^{22}$,	$3{\cdot}4\times10^{22}$,

Metal ...	Zn	Cd	Hg	C
	$5{\cdot}8\times10^{22}$,	$2{\cdot}7\times10^{21}$,	$4{\cdot}3\times10^{20}$,	$1{\cdot}08\times10^{19}$.

It is difficult to judge how much reliance ought to be placed on these magnetic methods of evaluating n until a more satisfactory explanation of the anomalous values of the Hall coefficient is forthcoming.

A detailed account of the modern experimental data relating to galvanomagnetic and thermomagnetic phenomena will be found in Baedeker's *Die Elektrischen Erscheinungen in Metallischen Leitern*, p. 94.

---

* *Phil. Mag.* vol. III. p. 643 (1902).

CHAPTER XVIII

THE EQUILIBRIUM THEORY OF ELECTRONIC CONDUCTORS

The Emission of Electrons from Hot Bodies and Thermoelectric Phenomena.

WHEN bodies are heated an emission of negative electrons is found to occur, which increases very rapidly with increasing temperature. The salient features of this phenomenon as they have revealed themselves experimentally may be briefly recapitulated as follows:

(1) The number of electrons* N emitted at different temperatures T is governed by the formula

$$N = AT^{\lambda} e^{-b/T} \quad \text{......................(I).}$$

A, λ and b are constants. A varies very much from one substance to another. λ is of the order unity, and its precise value makes very little difference to the formula. b in equivalent volts is always comparable with five.

(2) The emission (evaporation) of electrons is accompanied by an absorption of heat†. In the case of the only metals (osmium, tungsten and platinum) which have been examined the magnitude of this effect is about what would be expected from the values of b.

(3) The absorption‡ of electrons by a cold metal is accompanied by a liberation of heat. For a large number of metals this has been found to be approximately what the value of b would lead us to expect.

---

* O. W. Richardson, *Phil. Trans.* (A), vol. CCI. p. 543 (1903).

† Wehnelt and Jentzsch, *Ann. der Physik*, vol. XXVIII. p. 537 (1909); Cooke and Richardson, *Phil. Mag.* vol. XXV. p. 624 (1913), vol. XXVI. p. 472 (1913); We hne and Liebreich, *Verh. der deutsch. physik. Ges.* 15. Jahrgang, p. 1057 (1913).

‡ Richardson and Cooke, *Phil. Mag.* vol. XX. p. 173 (1910); vol. XXI. p. 40 1911).

(4) The energy* of, and the distribution of velocity among, the emitted electrons is identical with that given by Maxwell's law for molecules having the same mass as an electron at the same temperature as the emitting metal.

In addition to the points outlined above a large amount of information bearing on this subject in other ways has been accumulated. For an account of this the reader may refer to *The Electrical Properties of Flames and of Incandescent Solids*, by H. A. Wilson (University of London Press (1912)), or to an article by the writer in the *Handbuch der Radiophysik* (University of Leipzig Press (1914)).

In some cases, notably those furnished by the alkali metals†, this emission may be greatly augmented by the occurrence of chemical action. In fact, some recent writers have taken the standpoint that chemical action is an essential condition for the emission to occur. This is very unlikely on general grounds. For it amounts to denying that the simplest type of chemical action, namely, the decomposition of an elementary atom into a negative electron and a positive ion, can ever occur under the influence of heat alone. In the opinion of the writer, the facts also are decisively against it. The behaviour of platinum, which is the substance which has been most thoroughly studied, is very difficult to reconcile with any chemical theory, and the currents which can be obtained from incandescent tungsten are much too large to be attributed to chemical action‡.

We shall therefore assume that there is an emission of electrons from elementary and compound substances which is a purely thermal effect and try to find out what laws we should expect such a phenomenon to obey. It is convenient to have a separate name for an emission of this kind, and we shall describe the emission of electrons which occurs under the influence of heat by the term thermionic.

It is also worth while to consider for a moment what laws we should expect an emission to observe which is conditioned by chemical action. The electrons which are liberated are then to be regarded as one of the products, either intermediate or final, of the chemical action, and the extent to which they are formed will be determined by the laws which govern the formation of other chemical products. Now, the law which governs the rate of chemical actions in general is not essentially different from (I), and something very like the other results enumerated on p. 441 would also follow in this case, so that such considerations will not enable us to distinguish between the two phenomena. The only satisfactory criterion is whether the emission of electrons is accompanied, *pari passu*, by chemical combination or decomposition of the more generally recognized type. There is no evidence that this is universally the case.

---

* Richardson and Brown, *Phil. Mag.* vol. XVI. p. 353 (1908); Richardson, *Phil. Mag.* vol. XVI. p. 890 (1908); vol. XVIII. p. 681 (1909).

† Cf. Fredenhagen, *Verh. der deutsch. physik. Ges.* 14. Jahrgang, p. 384 (1912).

‡ Cf. O. W. Richardson, *Phil. Mag.* vol. XXVI. p. 345 (1913).

We shall now return to the consideration of the theory of the purely thermal emission. This phenomenon may be attributed to the increased kinetic energy of the electrons at high temperatures enabling them to overcome the forces which tend to retain them in the conductor. The rate of variation of the emission with the temperature may be calculated in a variety of ways. It will be conducive to clearness if we sacrifice generality to a slight extent and make a definite hypothesis about the structure of the interior of a metallic conductor: although many of our results will be more general than the hypothesis we are making. We have seen that the electrons in a good conductor behave as though they were acted upon by fixed centres of force varying as the inverse third power of the mutual distance. The potential energy of an electron must therefore be continually varying from point to point of its path. We shall suppose that the potential energy of an electron inside a metal is a function of its position only. In other words, we are supposing that the interior of a metal may be sufficiently described by mapping it out by a series of fixed surfaces, the level surfaces of W, the potential energy of an electron. These surfaces are supposed to be definite and characteristic for each conductor. The fact that the electrons are in motion prevents this from being a complete representation of the state of affairs, but it will be fairly certain to give a satisfactory account of the more important features of the phenomena.

The state of the electrons at any point of such a system as we are considering is defined by the average number ν of them per unit volume at that point at any instant, by the potential energy W of the electrons at that point, and by the temperature T, which is proportional to the average kinetic energy of the electrons. At points just outside any conductor W takes the constant value W_0. In the state of equilibrium the values of W_0 are different and characteristic for different conductors. This definition of W_0 is not sufficiently exact. At points very close to the conductor an electron is attracted towards it by its mirror image in the conductor. Thus W increases for some little distance away from the conductor and the points at which it approaches a sensibly constant value are not immediately outside the bounding surface. On the other hand, at considerable distances from the surface W will be affected by the potential of other bodies in the neighbourhood. The true value of W_0 is the value of W at points at a considerable distance from the surface, either in a cavity inside the conductor or outside if there are no other bodies in the neighbourhood.

It follows, as a result of our investigation of the kinetic theory of matter (p. 404), that at any point in a system in equilibrium at temperature T,

$$dn = \nu\, d\tau = Ke^{-W/RT}\, d\tau \quad \text{..................(1)},$$

where dn is the number of electrons, which participate in thermal phenomena, in the element of volume $d\tau$, and K is constant throughout the system, being a function of T only.

We shall now consider the formulae which determine the equilibrium between the external free electrons and the internal electrons which can become free. Confining our attention to a single conductor, consider first the special case in which there are a finite number p of finite internal regions each characterized by constant values ν_1, W_1, τ_1, ... ν_p, W_p, τ_p. Then if ν_0, W_0, τ_0 are the values of the corresponding variables just outside the conductor, we have

$$\frac{\nu_0}{e^{-W_0/RT}} = K = \frac{\nu_1}{e^{-W_1/RT}} = \dots = \frac{\nu_p}{e^{-W_p/RT}} = \frac{n}{\Sigma\tau e^{-W/RT}} \quad \text{......(2)},$$

where n is the total number of electrons which can become free that are present in the system. In general it will be possible to regard W as having a constant value only over a region of infinitesimal volume, so that

$$\frac{\nu_0}{e^{-W_0/RT}} = K = \frac{dn}{e^{-W/RT}\,d\tau} = \frac{n}{\int e^{-W/RT}\,d\tau} \quad \text{.........(3)},$$

where the integral is taken throughout the entire system. If we regard W_0 as a fixed constant,

$$\nu_0 = n \Big/ \int_\tau e^{-(W-W_0)/RT}\,d\tau \quad \text{..................(4)}.$$

The next step is to calculate the absorption of heat when electrons are allowed to evaporate from a conductor. Consider the conductor to be surrounded by an insulating boundary such that the volume of the space between the conductor and the boundary is v. If the boundary is displaced so that the volume increases by dv, work $p\,dv$ will be done by the equilibrium pressure, p, of the atmosphere of electrons. If we admit that such a change of the external volume is not accompanied by any material alteration of the mutual potential energy of the positively charged parts of the atoms which make up the conductor, the increment dS in the entropy of the system will be

$$dS = \frac{1}{T}(dU + p\,dv) \text{..............................(5)},$$

where

$$U = n\left(\tfrac{3}{2}RT + J\right) \text{..............................(6)},$$

$$J = \int We^{-W/RT}\,d\tau \Big/ \int e^{-W/RT}\,d\tau \quad \text{............(7)},$$

and

$$p = \nu_0 RT \quad \text{....................................(8)}.$$

n is the total number of electrons in the system which can become free, and the integrals in J are extended throughout the system. Since n is constant, it follows that when the volume is increased by dv the heat abstracted is

$$dQ = T\,dS = dv\left(n\frac{\partial J}{\partial v} + \nu_0 RT\right) \text{...............(9)}.$$

Since

$$J_{\tau+v+\delta v} = \int_{\tau+v+\delta v} We^{-W/RT}\,d\tau \Big/ \int_{\tau+v+\delta v} e^{-W/RT}\,d\tau$$

$$= \frac{\int_{\tau+v} We^{-W/RT}\,d\tau + \delta v\,We^{-W/RT}}{\int_{\tau+v} e^{-W/RT}\,d\tau + \delta v\,e^{-W/RT}},$$

$$\frac{\partial J}{\partial v} = \underset{\delta v = 0}{\mathrm{Lt}} \frac{1}{\delta v} \{J_{\tau+v+\delta v} - J_{\tau+v}\}$$

$$= \frac{e^{-W_0/RT}}{\int_{\tau+v} e^{-W/RT}\, d\tau} \left\{ W_0 - \frac{\int_{\tau+v} We^{-W/RT}\, d\tau}{\int_{\tau+v} e^{-W/RT}\, d\tau} \right\} \quad \ldots(10),$$

where $\int_{\tau+v}$ denotes that the integral is taken over the volume τ of the conductor plus the volume v of the external space.

Let n_v be the number of electrons in the external volume v and n_τ the number, of the kind contemplated, in the metal. Then in these problems n_v is always extremely small compared with n_τ. And since

$$\frac{n_\tau}{\int_\tau e^{-W/RT}\, d\tau} = K = \frac{n_v}{\int_v e^{-W/RT}\, d\tau}$$

by (3), we may replace

$$\int_{\tau+v} e^{-W/RT}\, d\tau \text{ in (10) by } \int_\tau e^{-W/RT}\, d\tau.$$

Also since

$$\frac{\int_v We^{-W/RT}\, d\tau}{\int_\tau We^{-W/RT}\, d\tau} = \frac{J_v}{J_\tau}\frac{n_v}{n_\tau} = \frac{J_v}{J_\tau}\frac{\frac{n_v}{v}}{\frac{n_\tau}{\tau}}\frac{v}{\tau} = \frac{J_v}{J_\tau}\frac{\nu_v}{\nu_\tau}\frac{v}{\tau},$$

and the following equation is true as regards order of magnitude,

$$\frac{\nu_v}{e^{-J_v/RT}} = K = \frac{\nu_\tau}{e^{-J_\tau/RT}},$$

the order of magnitude of

$$\frac{\int_v We^{-W/RT}\, d\tau}{\int_\tau We^{-W/RT}\, d\tau}$$

is that of $\frac{n_v}{n_\tau}$. Thus $\int_{\tau+v} We^{-W/RT}\, d\tau$ may also be replaced by

$\int_\tau We^{-W/RT}\,d\tau$ in (10). Whence

$$\frac{\partial J}{\partial v} = \frac{e^{-W_0/RT}}{\int_\tau e^{-W_0/RT}\,d\tau}(W_0 - J) \quad \ldots\ldots\ldots\ldots\ldots(11),$$

and from (3)

$$n\frac{\partial J}{\partial v} = \nu_0 (W_0 - J) \ldots\ldots\ldots\ldots\ldots\ldots\ldots(12).$$

By comparison with (9) the loss of heat which accompanies the escape of one electron is

$$\frac{dQ}{\nu_0 \delta v} = W_0 - J + RT = w + RT \quad \ldots\ldots\ldots\ldots(13).$$

Since RT is the external work, $w = W_0 - J$ is the internal latent heat of evaporation of one electron. Moreover, since S is a perfect differential,

$$\frac{\partial}{\partial v}\left(\frac{\partial S}{\partial T}\right)_v = \frac{\partial}{\partial T}\left(\frac{\partial S}{\partial v}\right)_T.$$

From (5) and (6)

$$\left(\frac{\partial S}{\partial T}\right)_v = \frac{n}{T}\left(\frac{3}{2}R + \frac{\partial J}{\partial T}\right) \text{ and } \left(\frac{\partial S}{\partial v}\right)_T = \frac{1}{T}\left(n\frac{\partial J}{\partial v} + p\right).$$

Thus $$n\frac{\partial J}{\partial v} + p = T\frac{\partial p}{\partial T} = T\left(\nu_0 R + RT\frac{\partial \nu_0}{\partial T}\right)$$

$$\left.\begin{aligned} &\text{and} && RT^2\frac{\partial \nu_0}{\partial T} = n\frac{\partial J}{\partial v} = \nu_0 w, \\ &\text{so that} && \nu_0 = Ae^{\int \frac{w}{RT^2}dT} = Ae^{-\frac{w}{RT} + \int \frac{1}{RT}\frac{\partial w}{\partial T}dT} \end{aligned}\right\} \ldots\ldots\ldots(14),$$

where A is a quantity, characteristic of the material, which is independent of T.

Equations (14) exhibit the relation between the equilibrium concentration ν_0 of the electrons in the external atmosphere, the temperature T and their internal latent heat of evaporation w. If w were a known function of T, (14) would enable us to completely specify ν_0 as a function of T. Some information about the way in which w depends on T may be obtained by the following slightly different thermodynamic argument.

The Relation with the Specific Heat of Electricity.

There is an intimate connection between ν_0, T, w and the specific heat of electricity σ within the conductor under consideration. Suppose that we have two conductors A and A' made of the same material, but maintained at the different absolute temperatures T and T'. They are to be of sufficiently great size and are connected by a thin conductor of the same material. The atmospheres of electrons about A and A' are enclosed and separated from each other by a suitable insulating boundary (see Fig. 50).

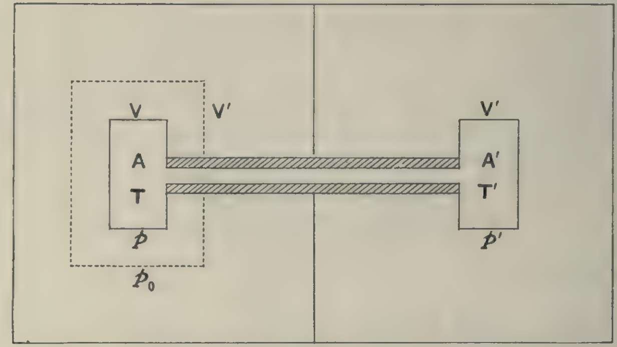


Fig. 50.

If contact difference of potential depends upon temperature, the potential V' at the surface of A' will not be the same as that at the surface of A. If the value at the latter surface is V, the electrons in the enclosure surrounding A will be at a different electric potential from those surrounding A'. Let eV' be $> eV$, and surround A with a screen which is permeable to the electrons and maintained at the potential V'. This may be imagined as a wire gauze of indefinite fineness which is connected to one end of a battery, the other end being connected to A. The electromotive force of the battery is $V' - V$. There is no work done by the battery, but even if there were it would not affect the argument. The nett effect of this arrangement is that the electrons in the enclosure outside the potential filter have the temperature T and the potential V'. Their pressure p_0 is different from the pressure p outside A, the relation being evidently

$$\log p_0 = \log p - \frac{e(V' - V)}{RT} \quad \text{..............(15).}$$

With the arrangement shown in the figure the electrons are made to perform a reversible thermodynamic cycle between the temperatures T and T' in the following manner:

By means of a suitable piston and cylinder arrangement N electrons are taken out of the enclosure bounding A. This operation is conducted at constant pressure p_0, the temperature being T and the potential V'. The external work done is NRT and the heat absorbed from A is $N\{w+e(V'-V)+RT\}$. The N electrons are then caused to expand adiabatically to the temperature T'. The work done during this adiabatic transformation is

$$\frac{NR}{1-\gamma}(T'-T),$$

where γ is the ratio of the specific heats of the electrons at constant pressure and at constant volume. The heat absorbed is nil and the pressure is changed from p_0 to

$$p_0\left(\frac{T}{T'}\right)^{\frac{\gamma}{1-\gamma}}.$$

The next step consists in expanding to the pressure p' at the constant temperature T'. The work done here is

$$-NRT'\log\frac{p'}{p_0}\left(\frac{T}{T'}\right)^{\frac{\gamma}{\gamma-1}},$$

and an equal amount of heat is absorbed. The electrons at the temperature T' and potential V' are then allowed to flow into A' under the uniform pressure p'. The work done during this stage is $-NRT'$ and the heat absorbed is $-N(w'+RT')$, where w' is the internal latent heat of evaporation of one electron at T'. Finally the electrons are allowed to flow down the connecting conductor back to A, when the whole system is in precisely the condition at which we started. There is no work done in the last operation and the heat absorbed is $Ne\int_{T'}^{T} SdT$, where S is the heat absorbed when unit quantity of electricity flows against unit rise of temperature.

Since the heat absorbed during the flow down the conductor is taken in at a series of continuously varying temperatures we have to apply the second law of thermodynamics in the integral form

$$\int \frac{dQ}{T} = 0.$$

This gives

$$\frac{w + e(V' - V) + RT}{T} - R \log \frac{p'}{p_0}\left(\frac{T}{T'}\right)^{\frac{\gamma}{1-\gamma}} - \frac{w' + RT'}{T'} + e\int_{T'}^{T} \frac{S}{T}\, dT = 0,$$

whence, by (15),

$$\frac{w}{T} - \frac{w'}{T'} + R\{\log p - \log p'\} - \frac{\gamma}{\gamma - 1}(\log T - \log T') + e\int_{T'}^{T} \frac{S}{T}\, dT = 0 \qquad \text{.........(16)},$$

or

$$\log p - \frac{\gamma}{\gamma - 1} \log T + \frac{w}{RT} + \frac{e}{R}\int^{T} \frac{S}{T}\, dT = A \qquad \text{......(17)},$$

and

$$\nu_0 RT = p = A_1 T^{\frac{\gamma}{\gamma-1}} \epsilon^{-\frac{w}{RT} - \frac{e}{R}\int_0^T \frac{S}{T} dT} \qquad \text{.........(18)},$$

or

$$\nu_0 = A_2 \epsilon^{-\frac{w}{RT} + \int^T \left(\frac{1}{\gamma - 1} - \frac{eS}{R}\right) \frac{dT}{T}} \qquad \text{...........(19)}.$$

In these equations A, A_1 and A_2 are constants which are characteristic of the substance and are *independent of* T. ϵ is used for the base of the natural logarithms to distinguish it from the electronic charge e.

A certain amount of caution is necessary in the interpretation of these equations. In measuring the specific heat of electricity σ, a continuous current is driven along by means of an impressed electromotive force. If we are to be certain that S and σ are the same thing it is desirable that the flow down the conductor should be a continuous process. This can be realized to any desired degree of approximation by making the cylinders sufficiently small and numerous and working them rapidly enough. As we have not proved that the thermal effects arising from a current inspired by diffusion are identical with those due to a current driven by an impressed electromotive force, it is desirable that the current in the unequally heated conductor should be driven in this way. Our cyclic process may be adapted to this end by introducing a battery in the wire, which is made indefinitely thin, whilst the conductors A and A' are indefinitely large. The transference in the wire is thus electrically driven and outside of A and A' it is effected mechanically, the whole process being continuous.

Under these circumstances the w of equations (16)—(19) is not necessarily the same thing as the w of equation (14). The latter refers to a virtual displacement of electrons across the boundary subject to the equilibrium conditions, whereas the former refers to a continuous stream such as actually occurs under an impressed force. Since the electric transference produces no permanent change in the materials, the difference between these two quantities will be equal to the difference in the quantities of energy which accompany the transference of unit quantity of electricity in the respective media on the two sides of the boundary, if such a difference exists. If we keep the symbol w for the w of equations (16)—(19) and denote the true latent heat of evaporation—the w of equation (14)—by ϕ we have

$$w = \phi - (\lambda_1 - \lambda_0),$$

where λ_1 is the energy transferred by unit electric current in the metal in unit time and λ_0 is the corresponding quantity for the current outside the metal. Thus instead of (19) we get

$$\nu_0 = A_2 \epsilon^{-\frac{1}{RT}\{\phi - (\lambda_1 - \lambda_0)\} + \int^T \left(\frac{1}{\gamma - 1} - \frac{e\sigma}{R}\right)\frac{dT}{T}} \quad \text{......(20)}.$$

We have now arrived at two different expressions (14) and (20) for ν_0 in terms of ϕ, σ and T. Since, if ϕ and σ could be expressed in terms of T, these expressions would have to become identically equal, it follows that

$$\frac{\lambda_1 - \lambda_0}{RT} + \int^T \left(\frac{1}{\gamma - 1} - \frac{e\sigma}{R}\right)\frac{dT}{T} = \int^T \frac{1}{RT}\frac{d\phi}{dT}\,dT$$

for all values of T. Hence σ will be given by each of the equations:

$$\sigma = \frac{1}{e}\left\{\frac{R}{\gamma - 1} + T\frac{d}{dT}\left(\frac{\phi - w}{T}\right) - \frac{d\phi}{dT}\right\} \quad \text{.........(21)},$$

$$= \frac{1}{e}\left\{\frac{R}{\gamma - 1} + \frac{w - \phi}{T} - \frac{dw}{dT}\right\} \quad \text{.....................(22)},$$

$$= \frac{1}{e}\left\{\frac{R}{\gamma - 1} + T\frac{d}{dT}\frac{\lambda_1 - \lambda_0}{T} - \frac{d\phi}{dT}\right\} \quad \text{...........(23)},$$

$$= \frac{1}{e}\left\{\frac{R}{\gamma - 1} - \frac{\lambda_1 - \lambda_0}{T} - \frac{dw}{dT}\right\} \quad \text{..................(24)}.$$

Approximation to v_0 as a function of T.

If the electrons are treated as point charges, they have three degrees of freedom and $\gamma = \frac{5}{3}$. If the collisions of the electrons occur with centres of force varying as a power of the mutual distance, it follows from the kinetic theory calculations in the last chapter that $\frac{d}{dT}\left(\frac{\lambda_1 - \lambda_0}{T}\right) = 0$, so that, under these conditions,

$$\sigma = \frac{1}{e}\left\{\frac{3}{2}R - \frac{\partial\phi}{\partial T}\right\} \quad\text{.....................(25).}$$

The value of σ is different for different metals and may be either positive or negative. In general it also varies with the temperature, often in a complicated manner. If the thermoelectric powers of different pairs of metals were accurately linear functions of the

Metal	Temperature ° C.	σ Erg/El. mag. C.G.S. unit	Observer
Mercury	+ 50	$6{\cdot}8 \times 10^2$	Schoute
	+100	$8{\cdot}6 \times 10^2$	"
	+150	$10{\cdot}6 \times 10^2$	"
Copper	−100	$0{\cdot}9 \times 10^2$	Berg
	0	$1{\cdot}6 \times 10^2$	"
	+100	$2{\cdot}0 \times 10^2$	"
	+300	$2{\cdot}1 \times 10^2$	Lecher
	+500	$2{\cdot}6 \times 10^2$	"
Silver	+100	$3{\cdot}46 \times 10^2$	"
	+300	$4{\cdot}20 \times 10^2$	"
	+500	$4{\cdot}95 \times 10^2$	"
Platinum	− 50	$-9{\cdot}4 \times 10^2$	Berg
	0	$-9{\cdot}1 \times 10^2$	"
	+ 50	$-9{\cdot}0 \times 10^2$	"
	+100	$-9{\cdot}1 \times 10^2$	"
Iron	0	$-4{\cdot}0 \times 10^2$	"
	+100	$-12{\cdot}4 \times 10^2$	"
	+100	$-13{\cdot}8 \times 10^2$	Lecher
	+200	$-16{\cdot}8 \times 10^2$	"
	+300	$-14{\cdot}2 \times 10^2$	"
	+400	$-7{\cdot}5 \times 10^2$	"
Constantan	0	$-23{\cdot}0 \times 10^2$	"
	+200	$-19{\cdot}9 \times 10^2$	"
	+400	$-13{\cdot}7 \times 10^2$	"

temperature, then the difference in the values of σ for any two metals would be proportional to the absolute temperature (equations (51), p. 426). This is not borne out by the preceding experimental values of σ which are taken from a table compiled by Baedeker*. Thus for iron σ is negative and its numerical value is greatest at just over 200° C. It is questionable, however, whether the reliability of the experimental determinations of $\frac{\partial E}{\partial T}$ and σ warrants us in rejecting equation (50), p. 426, Chap. XVII.

In all cases it has been found that $e\sigma$ is a good deal smaller than $\frac{3}{2}R$, so that $\frac{3}{2}R$ may be taken as a first approximation to the value of $\frac{d\phi}{dT}$. Thus neglecting the second and higher derivatives of ϕ we have, approximately,

$$\phi = \phi_0 + \tfrac{3}{2}RT \qquad \text{.......................(26)},$$

where ϕ_0 is independent of T. Substituting this value of ϕ for w in (14) we obtain

$$\nu_0 = A_1 T^{\frac{3}{2}} e^{-\frac{\phi_0}{RT}} \qquad \text{.......................(27)}.$$

In this equation A_1 and ϕ_0 are independent of T to the extent to which the assumed approximations are valid.

Now there is a simple relation between ν_0 and the number of electrons emitted by unit area of the conductor in unit time, provided we can neglect the reflexion, at the surface of the conductor, of the electrons which return to it from the exterior. It is known from experimental results that such reflexion is appreciable, but we shall disregard it for the present, and consider how it may be allowed for later. In that case, since ν_0 represents the number of electrons per cubic centimetre of the space in the steady state, the number emitted by the surface in unit time will evidently be equal to the number returned from the exterior in the same interval. This is

$$\begin{aligned} N &= \nu_0 \int_0^\infty du \int_{-\infty}^\infty dv \int_{-\infty}^\infty dw \; n \left(\frac{km}{\pi}\right)^{\frac{3}{2}} \epsilon^{-km(u^2+v^2+w^2)} \, du\,dv\,dw \\ &= \frac{\nu_0}{2} \frac{1}{(\pi k m)^{\frac{1}{2}}} = \nu_0 \left(\frac{R}{2\pi m}\right)^{\frac{1}{2}} T^{\frac{1}{2}} \qquad \text{.............................(28)}. \end{aligned}$$

---

* *Elek. Erscheinungen in Metallischen Leitern*, p. 76.

If we apply a sufficiently great potential difference all the N electrons emitted by the conductor may be collected on a neighbouring electrode. The ensuing saturation current is, per unit area of the emitting body,

$$Ne = A_1 \left(\frac{R}{2\pi m}\right)^{\frac{1}{2}} T^2 \epsilon^{-\phi_0/RT} = A_2 T^2 \epsilon^{-\phi_0/RT} \quad \text{......(29)},$$

where A_2 and ϕ_0 are constants characteristic of the substance but independent of T. Had we not made the approximations above the index of T would be slightly different from 2 and we should have had exponential factors of the type $\epsilon^{a_p T^p}$, where a_p is an extremely small constant. The values of these factors cannot differ much from unity. We should therefore expect (29) to give a fair representation of the mode of variation with temperature of the electrons emitted by hot bodies. The number of substances which have been tested in this way is now very considerable and for all of them the emission has been found to be consistent with an equation of the form $Ne = AT^{\frac{1}{2}}\epsilon^{-b/T}$. As, however, the variation is almost all in the exponential term the results can be fitted as well by $Ne = AT^2\epsilon^{-b/T}$, by taking a slightly different value of the constant b in the exponential index.

We shall now consider the bearing of these effects on the nature of contact electromotive force.

Contact Difference of Potential.

Imagine an enclosure limited by an insulating boundary, maintained at temperature T. Suppose the enclosure to contain q material bodies, arranged in any manner, and that the whole system has come to a state of thermal equilibrium. The temperature will then be uniform and equal to T throughout the system. In general the surfaces of the bodies will assume different potentials $V_1, V_2, \dots V_q$. Clearly

$$e(V_m - V_p) = W_0^m - W_0^p \quad \text{..................(30)},$$

where W_0^m is the potential energy of an electron just outside the mth body. If we can show that $W_0^m - W_0^p$ are uniquely determined by the constitutive equations of the system it will follow that $V_m - V_p$ represent true intrinsic differences of potential. Let ν_0'

etc. represent the concentrations of the electrons at points just outside the different bodies. Since there is equilibrium between the electrons which are just outside each pair of bodies we evidently have $q-1$ equations of the type

$$\nu_0{}^m/\nu_0{}^p = \epsilon^{-\frac{W_0{}^m - W_0{}^p}{RT}}.$$

From (14) and (12) we see that there are q equations of the type $\nu_0{}^m = \phi_m(W_0{}^m J_m A_m T)$. From (7) there are also q equations of the type $J_m = \psi_m(T)$ and evidently there are q equations of the type $A_m =$ constant. Altogether there are $4q-1$ equations between the $3q$ variables ν_0, J and A, the $q-1$ differences $W_0{}^m - W_0{}^p$ and T. Thus there are $4q-1$ equations and $4q$ variables, so that if one of the latter, say T, is given all the variables including the $q-1$ differences $W_0{}^m - W_0{}^p$ are determined. These equations involve neither the size, shape nor relative orientation of the bodies, so that the differences of potential $V_m - V_p$ are characteristic of the substances under consideration. They are clearly the same whether the bodies are in contact or not.

Since from (13) $w = W_0 - J$ it follows that

$$W_0{}^m - W_0{}^p = w_m - w_p + J_m - J_p.$$

We shall see below (p. 457) that $J_m - J_p$ is small compared with $w_m - w_p$; so that according to this view the contact difference of potential may be estimated from the internal latent heat of evaporation of the electrons. The experimental determinations of w for different substances are not yet sufficiently trustworthy to furnish a satisfactory test of this relation; although such indications as there are tend to show that w is smaller for electropositive than for electronegative elements.

The Peltier Effect.

We shall now proceed to obtain expressions for the various physical quantities which are grouped under thermoelectric phenomena. Let us first consider the Peltier coefficient P. To do this we need only to consider a reversible isothermal cycle involving two different conductors. The arrangement and operation is in fact much the same as that shown in Fig. 50 except that the conductors A and A' are of different materials and are at the same temperature T. Let the suffixes 1 and 2 be used to denote the various physical quantities which relate to the separate materials, the notation being otherwise as before. Let eV_2 be $> eV_1$. Surround the first conductor, as before, by a potential filter maintained at V_2. This will reduce the pressure of the electrons from p_1 to p_1^0 and will change their potential energy from eV_1 to eV_2 but will not affect their temperature T. The cycle commences by removing N electrons from the enclosure surrounding A_1 under the uniform pressure p_1^0. The work done is NRT and the heat absorbed is, by virtue of (13), $N[w_1 + e(V_2 - V_1) + RT]$. Next expand at T from p_1^0 to p_2. The work done and heat absorbed are each equal to $NRT \log (p_1^0/p_2)$. Next condense at p_2 and V_2 in the second body. The work done is $-NRT$ and the heat absorbed is $-N[w_2 + RT]$. The electrons are then allowed to flow along the connecting conductor to A_1, thus completing the reversible cycle. During this operation no work is done, but there is an absorption of heat NQe at the junction. The total amount of work in this isothermal reversible cycle is $NRT \log (p_1^0/p_2)$. Since the cycle is isothermal this must vanish, so that $p_1^0 = p_2$. The total absorption of heat is $N[w_1 + e(V_2 - V_1) + RT + RT \log (p_1^0/p_2) - w_2 - RT + Qe]$, and since this also must vanish we have

$$Qe = w_2 - w_1 - e(V_2 - V_1) \quad \text{..............(31)}.$$

If Q is to be the same thing as the coefficient P which is measured in experiments on the Peltier effect, it is necessary that the cycle should be operated under the same conditions as when a current flows continuously under an impressed electromotive force. Hence, as in dealing with the Thomson effect, w_2 and w_1 will be the values of the latent heats which correspond to a steady flow and not to equilibrium conditions. Denoting the former by w and the latter by ϕ as before, we therefore have the equations

$$eP = w_2 - w_1 - e(V_2 - V_1) \quad \text{......................(32)},$$

$$= \phi_2 - \phi_1 - (\lambda_2 - \lambda_1) + e(V_2 - V_1) \quad \text{.........(33)},$$

$$= J_1 - J_2 - (\lambda_2 - \lambda_1) \quad \text{..........................(34)},$$

since $\phi \equiv W_0 - J$.

A series of experimental values of P as well as the values of the thermoelectric power of the same pairs of metals at the same temperatures are given in the following table, taken from Baedeker*.

Materials	Temperature C.	Peltier coefficient in		Thermoelectric† power. $1=10^{-6}$ volt per degree C.
		Millicalories per coulomb	Ergs per E.M. unit	
Cu → Ag	0	0·137	$0{\cdot}0576\times10^6$	2·12
Fe → Constantan	0	3·4	$1{\cdot}42\times10^6$	(47·7)
	100	4·1	$1{\cdot}72\times10^6$	(50·7)
	200	5·5	$2{\cdot}31\times10^6$	(53·7)
Pb → Constantan	0	1·90	$0{\cdot}80\times10^6$	27·1
	100	2·73	$1{\cdot}14\times10^6$	33·5
	200	3·6	$1{\cdot}51\times10^6$	39·9
	300	4·4	$1{\cdot}84\times10^6$	46·3
Cd → Pb	0	0·197	$0{\cdot}082\times10^6$	3·03
	100	0·390	$0{\cdot}163\times10^6$	4·48
	200	0·646	$0{\cdot}271\times10^6$	5·93
Cu → Ni	0	1·9	$0{\cdot}80\times10^6$	(27)
	100	2·2	$0{\cdot}92\times10^6$	(30)
	about 220	2·5 (max.)	$1{\cdot}05\times10^6$	(34)
	about 340	1·9 (min.)	$0{\cdot}80\times10^6$	(23)
	450	2·4	$1{\cdot}00\times10^6$	(25)

In all these cases the Peltier coefficient when expressed in equivalent volts is of the order 10^{-2}—10^{-3} volt. The ratio of the thermal to the electrical conductivity of metals, which contains λ as a factor, shows that $\frac{1}{e}(\lambda_2-\lambda_1)$ cannot be much greater numerically than this value of P. On the other hand V_2-V_1 may amount to several volts in extreme cases. It follows that $e(V_2-V_1)$, w_2-w_1 and $\phi_2-\phi_1$ are approximately equal to one another. Thus J_m-J_p is small compared with w_m-w_p as stated on p. 455.

---

* *Loc. cit.* p. 73.

† The values in brackets were obtained from specimens of material different from those used in measuring the Peltier coefficient.

The Thomson Effect.

A number of expressions for σ have already been given (p. 451). For the purpose of comparing the results of different theories it is desirable to express some of them in terms of the quantities J etc. which describe the internal structure of the material instead of ϕ etc. Consider, for example, equation (23). We have $\frac{\partial \phi}{\partial T} = \frac{\partial}{\partial T}(W_0 - J)$ and from (2) and (14)

$$\frac{\partial \log \nu_0}{\partial T} = \frac{\partial \log K}{\partial T} + \frac{W_0}{RT^2} - \frac{1}{RT}\frac{\partial W_0}{\partial T} = \frac{W_0}{RT^2} - \frac{J}{RT^2}.$$

Thus

$$\frac{\partial \phi}{\partial T} = \frac{J}{T} - \frac{\partial J}{\partial T} + RT\frac{\partial \log K}{\partial T} = -T\frac{\partial}{\partial T}\left\{\frac{J}{T} - R\log K\right\},$$

so that

$$\sigma = \frac{1}{e}\left\{\frac{R}{\gamma - 1} + T\frac{\partial}{\partial T}\left(\frac{J + \lambda_1 - \lambda_0}{T} - R\log K\right)\right\} \quad \ldots(35).$$

This expression is more general than equation (61) of the last chapter, and the two become identical only in rather simple cases. We have seen that the kinetic theory methods employed in Chapter XVII can only be regarded as strictly accurate when the linear dimensions of the regions, within which the forces exerted on the electrons are appreciable, are small compared with the distance between collisions. In that case the potential energy w of the electrons will only differ by a negligible amount from the mean value J. Also the A of the last chapter is equal to $N\left(\frac{h}{\pi}\right)^{\frac{3}{2}}$ if N is the number of free electrons in unit volume. Since

$$\log N = \log K - \frac{w}{RT} = \log K - \frac{J}{RT},$$

$$\frac{\partial}{\partial T}\left\{\frac{J}{T} - R\log K\right\} = -R\frac{\partial}{\partial T}(\log N) = -R\frac{\partial}{\partial T}(\log A) - \frac{3}{2}\frac{R}{T}.$$

Thus, from (35),

$$\sigma = \frac{R}{e}\left\{\frac{1}{\gamma - 1} - \frac{3}{2} - T\frac{\partial}{\partial T}\left(\log A_1 - \frac{\lambda_1 - \lambda_0}{T}\right)\right\}$$

$$= -\frac{R}{e}T\frac{\partial}{\partial T}\left\{\log A_1 - \frac{\lambda_1 - \lambda_0}{T}\right\}, \text{ if } \gamma = \frac{5}{3}.$$

This is in agreement with (61), since under the conditions contemplated $\frac{\partial}{\partial T}\left(\frac{\lambda_1 - \lambda_0}{T}\right) = -\frac{\partial \mu_1}{\partial T}$.

These equations, which have been deduced from thermodynamic considerations, are thus in agreement with those given by the kinetic theory calculations when the latter are accurate. In virtue of (34), equation (35) satisfies the condition

$$\sigma_1 - \sigma_2 = T\frac{\partial}{\partial T}\left(\frac{P}{T}\right)$$

postulated by Lord Kelvin's Theory.

Thermoelectromotive Force.

Having found expressions for the Peltier and Thomson coefficients we can, by means of the energy equation (48) of the last chapter, deduce from them the value of the thermoelectromotive force of the complete circuit. For a circuit of two materials with junctions at T' and T_0 one finds in this way, for example,

$$E = (V_1 - V_2)_{T'} - (V_1 - V_2)_T + \frac{1}{e}\int_{T_0}^{T'} \frac{w_2 - w_1 - \phi_2 + \phi_1}{T}\, dT \quad \ldots(36).$$

In a similar way equations may be obtained containing other combinations of the variables which occur in equations (21)—(24) and (32)—(34).

By making the assumption of thermodynamic reversibility and substituting the values of P and σ from equations (23) and (32) in equation (49), p. 426, we obtain

$$\left|\frac{w_1 - w_2 + e(V_2 - V_1)}{T}\right|_{T_0}^{T'} + \int_{T_0}^{T'}\left\{\frac{w_1 - w_2 - \phi_1 + \phi_2}{T^2} - \frac{1}{T}\frac{\partial}{\partial T}(w_1 - w_2)\right\} dT = 0,$$

and, after integrating by parts,

$$\left|\frac{e(V_2 - V_1)}{T}\right|_{T_0}^{T'} = \int_{T_0}^{T'} \frac{\phi_1 - \phi_2}{T^2}\, dT,$$

or

$$\phi_1 - \phi_2 = e\left\{V_1 - V_2 - T\frac{\partial}{\partial T}(V_1 - V_2)\right\} \quad \ldots\ldots\ldots(37).$$

By making use of (37) the following expressions for E may be obtained from (36) or the equivalent equations:

$$E = \frac{1}{e}\int_{T_0}^{T'} \frac{w_2 - w_1 + e(V_1 - V_2)}{T}\, dT \quad \text{.....................(38)}$$

$$= \frac{1}{e}\int_{T_0}^{T'} \frac{\phi_2 - \phi_1 - (\lambda_2 - \lambda_1) + e(V_1 - V_2)}{T}\, dT \quad \text{......(39)}$$

$$= \Big| V_1 - V_2 \Big|_{T_0}^{T'} + \frac{1}{e}\int_{T_0}^{T'} \frac{w_2 - w_1 - \phi_2 + \phi_1}{T}\, dT \quad \text{......(40)}$$

$$= \Big| V_1 - V_2 \Big|_{T_0}^{T'} + \frac{1}{e}\int_{T_0}^{T'} \frac{\lambda_1 - \lambda_2}{T}\, dT \quad \text{..................(41).}$$

The values in terms of J_1, J_2, etc. may be derived by obvious substitutions. The thermoelectric power $\frac{\partial E}{\partial T'}$ is given by differentiating (38)—(41) respectively by the upper limit T'. These equations obviously satisfy Lord Kelvin's condition

$$\frac{\partial E}{\partial T'} = \frac{P}{T}.$$

Alternative Expressions for the Thermoelectric Coefficients.

If we wish to express these quantities in terms of the number of electrons $\left(\nu = \frac{dn}{d\tau}\right)$ in unit volume at each point of the conductor, instead of in terms of the functions J which depend on their mean potential energy, we may proceed as follows:

Since from (3) $\nu = Ke^{-\frac{W}{RT}}$, we have

$$\left.\begin{aligned} J_1 &= \int_{\tau_1} \frac{\nu}{K}\{RT\log K - RT\log\nu\}\, d\tau \div \int \frac{\nu}{K}\, d\tau \\ &= RT\log K - \frac{RT}{n_1}\int_{\tau_1} \nu\log\nu\, d\tau \\ &= RT\log K - \frac{RT}{n_1}\int_{n_1}\left\{\log\frac{dn}{d\tau}\right\} dn \end{aligned}\right\} \text{......(42).}$$

Thus, for example,

$$eP = J_1 - J_2 + \lambda_1 - \lambda_2$$

$$= RT\left\{\frac{\lambda_1}{RT} - \frac{1}{n_1}\int_{n_1}\left\{\log\frac{dn}{d\tau}\right\} dn - \left[\frac{\lambda_2}{RT} - \frac{1}{n_2}\int_{n_2}\left\{\log\frac{dn}{d\tau}\right\} dn\right]\right\},$$

and similarly with the other quantities σ_1, E and $\frac{\partial E}{\partial T}$.

If the distance between collisions is large compared with the distances within which the forces are appreciable, then the potential energy of the electrons is constant in all but a small fraction of the available space. In this case we can replace the average values of ν and W by actual values without serious error. If, in addition, the law of force during collisions varies as a power of the distance, we see from the last chapter that $\frac{\partial}{\partial T}\left(\frac{\lambda}{T}\right) = 0$. And, finally, if the law of force is the same for the different materials then, for any pair, $\lambda_m - \lambda_p = 0$. Under these circumstances the expressions which have been found reduce to

$$Pe = RT \log \nu_2/\nu_1 \qquad \text{(43)},$$

$$\sigma_1 = \frac{R}{e}\left[\frac{1}{\gamma - 1} - T\frac{\partial}{\partial T}(\log \nu_1)\right] \qquad \text{(44)},$$

$$E = \frac{R}{e}\int_{T_0}^{T} \log\left(\frac{\nu_2}{\nu_1}\right) dT \qquad \text{(45)},$$

$$\frac{\partial E}{\partial T} = \frac{R}{e}\log\frac{\nu_2}{\nu_1} \qquad \text{(46)}.$$

These equations are exact if the collisions are always like those between hard elastic spheres.

Comparison with Experiment.

There are a number of reasons why formulae (43)—(46) are unsatisfactory. In the first place the ratio of thermal and electrical conductivities indicates that the force during collisions does not vary very sharply with the mutual distance. Moreover the computed strength of the centres indicates that the electrons are never free from very considerable forces. Thus the theoretical conditions are far from being satisfied. In the second place the formulae are contradicted by the thermoelectric data themselves. By hypothesis all the electrons are to be treated alike, so that ν in equations (43)—(46) represents the number of current-carrying electrons in unit volume of the material. Now the electrical conductivity is proportional to ν, and there is no reason to expect that the other factors which enter into it will not be of the same order of magnitude for different substances. We should therefore expect that ν_2/ν_1 would, for different substances, be comparable with the ratio of their specific electrical conductivities. Formula (43) would therefore lead us to expect an exceptionally large Peltier effect at the junction between a very good and a very bad conductor. The Peltier effect in a number of cases of this kind has recently been investigated by Koenigsberger and Weiss*. Although these experiments are difficult to make, they seem to have established that the Peltier effect in such cases is not, in general, of exceptional magnitude. In some cases it was found to be in the opposite direction to that given by equation (43). The more complete theory involves two considerations which are neglected in equation (43). In the first place, the law of force during collisions may vary from one material to another. It is, however, extremely improbable that the term $\lambda_1 - \lambda_2$, which arises from this circumstance, can be comparable with the term involving $\log \nu_1/\nu_2$, when the ratio of the two conductivities is very great. We are thus driven to the conclusion that the difference in the potential energy of the electrons is not, in general, measured, even approximately, by $RT \log \nu_1/\nu_2$, where ν_1 and ν_2 are the volume concentrations of the free electrons. This objection cannot be made in the case of the formulae given by the more complete theory, which make the Peltier effect depend on the mean potential energy of all the electrons which may, from time to time, become free under the dynamical actions actually occurring, and not on the actual or average number of those free instantaneously. The number of the former may be quite considerable, although, on account of the intensity of the attraction of the rest of the atom, very few of them are able to get free enough to take part in carrying the current at any instant.

A second point which is of interest in this connection is that raised by the phenomena exhibited at the melting-point. For all the metals which have been tried, except antimony and bismuth, the specific conductivity of the solid at the melting-point is greater than that of the liquid. The changes are quite considerable, the ratio of the two specific conductivities varying from 1·34 in the case of sodium to 4·1 in the case of mercury. In the case of antimony the ratio is 0·70 and in the case of bismuth 0·46. All the metals except antimony and bismuth contract when they solidify. In accordance with (43) and (46) we should expect these changes in the electrical conductivity to be accompanied by very considerable changes in the Peltier effect and the thermoelectric power. Such effects have been looked for most carefully and with negative results. It is clear that the change in the thermoelectric power, if there is any, is much smaller than we should expect from (46). However, it cannot be said that the absence of Peltier effect at the melting-point is properly understood on the more general theory. It is evident that it might happen in a number of ways which it would take too long to discuss here.

---

* *Ann. der Physik*, vol. xxxv. p. 1 (1911).

In the case of the iodiferous cuprous iodide, to which reference has already been made in connection with the Hall effect, Baedeker* finds equation (46) to be accurately verified if the relative values of ν are measured by the relative conductivities. It seems as though the conditions affecting the motion of the electrons are much simpler in the case of this substance than in the case of the metals.

The Conducting Electrons.

The expressions which have been obtained for the various thermoelectric quantities involve integrals which are extended over all the electrons which are able at any time to take part in thermoelectric phenomena. In general we should expect only part of these electrons to be capable of engaging in the transportation of the electric current. Many of them will be so strongly attracted to the atom nearest to them that they will only rarely be able to escape from its immediate neighbourhood. Such electrons will only be slightly displaced under the influence of an external field and will not participate in the conveyance of the conduction current.

It seems natural to suppose that the conduction current is carried by those electrons of the class contemplated which are instantaneously executing open, as opposed to closed, orbits. An expression for the number N of electrons which are executing open orbits may be obtained on the assumption, which accords with the results of the discussion in the last chapter, that the electrons are attracted by centres of force varying inversely as the cube of the mutual distance. Let R be the instantaneous radius to the centre and θ the angle this makes with the direction of R when $t=0$. Let $\dot{T}$ be the tangential and $\dot{R}$ the radial component of velocity. Let V_1 be the velocity from infinity, then

---

* *Nernst Festschrift*, p. 62, Halle (1912).

$$\tfrac{1}{2} m V_1^2 = - W.$$

It is shown in Routh's *Dynamics of a Particle* (Cambridge, 1898), p. 231, that all the orbits for which $V_1^2 < \dot{T}^2$ are open.

The equation to the others is given by

$$\frac{1}{R} = A e^{n\theta} + B e^{-n\theta},$$

where $\dot{T}_0 = h(A+B)$, $\dot{R}_0 = nh(B-A)$, $\dot{T}_0$ and $\dot{R}_0$ are the values of $\dot{T}$ and $\dot{R}$ when t and θ are zero, $n^2 = \frac{V_1^2}{\dot{T}_0^2} - 1$ and $h = R\dot{T}$. Evidently R can only become infinite if A and B have opposite signs. Since

$$B = - A \frac{1 + n\frac{\dot{T}_0}{\dot{R}_0}}{1 - n\frac{\dot{T}_0}{\dot{R}_0}},$$

the condition that B and A should have the same sign is

$$n^2 \frac{\dot{T}_0^2}{\dot{R}_0^2} = \frac{V_1^2 - \dot{T}_0^2}{\dot{R}_0^2} > 1,$$

i.e.

$$V_1^2 > \dot{T}_0^2 + \dot{R}_0^2.$$

The complete condition for an open orbit is therefore that the initial kinetic energy should exceed the kinetic energy due to the attractive forces supposed to act from infinity to the point considered. If we put $\phi = - W = \frac{1}{2} m V_1^2$, we evidently have

$$\left.\begin{aligned} \delta N &= \iiint \delta n \left(\frac{km}{\pi}\right)^{\frac{3}{2}} e^{-km(u^2+v^2+w^2)}\, du\, dv\, dw \\ &= \delta n \left(\frac{km}{\pi}\right)^{\frac{3}{2}} \int_{V_1}^{\infty} dr \int_0^{\pi} d\theta \int_0^{2\pi} d\phi\, r^2 \sin\theta\, e^{-kmr^2} \\ &= \delta n \frac{2}{\sqrt{\pi}} \left\{ \sqrt{\frac{\phi}{RT}}\, e^{-\phi/RT} + \int_{\left(\frac{\phi}{RT}\right)^{\frac{1}{2}}}^{\infty} e^{-x^2} dx \right\} \\ &= \delta n \chi \left(\frac{W}{RT}\right) \end{aligned}\right\} \quad \ldots\ldots(47),$$

where δn denotes the number of electrons of all sorts at a point where their potential energy is W at any instant, and δN denotes the number of them which will execute open orbits. This calculation assumes that the distribution of kinetic energy among the particles is independent of the potential energy, a conclusion which we have already established (see p. 405).

Now the number dn of electrons of all speeds in the element of volume $d\tau$ is equal to $\nu d\tau$ where, from (3),

$$\nu = \frac{ne^{-W/RT}}{\int_\tau e^{-W/RT}\,d\tau} = Ke^{-W/RT},$$

whence the total number N of current-carrying electrons in the volume τ is

$$\begin{aligned} N &= \int_\tau \nu\chi(W/RT)\,d\tau \\ &= K\int_\tau \chi\left(\frac{W}{RT}\right) e^{-W/RT}\,d\tau \\ &= n\int_\tau \chi\left(\frac{W}{RT}\right) e^{-W/RT}\,d\tau \Big/ \int_\tau e^{-W/RT}\,d\tau \quad \text{.........(48).} \end{aligned}$$

It is impossible to predict anything very precise about the behaviour of integrals like those in (48); but it will be observed by comparing with formula (39), for example, and neglecting $\lambda_2 - \lambda_1$, that both the number of current-carrying electrons in a cubic centimetre and the thermoelectric power are functions of the distribution of $\frac{W}{RT}$ in the space inside the conductor. We should therefore expect that any cause which produced a *continuous* alteration of the electric conductivity of the substance at a given temperature would produce a corresponding alteration in the thermoelectric power. The changes in the one quantity would be expected to follow changes of the other quantity in a corresponding manner.

Evidence of a correspondence of this kind exists in the case of alloys. Their electrical behaviour is found to depend very largely on the intimate structure of the alloy. From this point of view alloys may be regarded as falling into two main classes: (1) those which consist of mixtures of crystals of the constituents and (2) those formed of crystals whose constituents are mutually soluble. In some cases the constituents are only able to dissolve in one another to a definite limited extent. In other cases chemical compounds are formed which give rise to considerable complication. The behaviour in such cases, however, is determined by the mutual solubility of the compounds and the independent constituents; so that the effect of chemical combination is practically to increase the number of substances which have to be considered. Otherwise their behaviour is similar to that of cases where no chemical combination appears to occur.

Let us consider the two simple cases of complete immiscibility and of complete mutual solubility, from which all the others may be developed. In the case where the constituents are entirely immiscible the specific electrical conductivity, expressed as a function of the composition, changes linearly from the value characteristic of one of the pure substances to that characteristic of the other. Precisely the same statement is true of the thermoelectric power referred to a standard metal. The behaviour of alloys whose constituents are mutually soluble is entirely different in character. The addition of a small quantity of either constituent to the other pure metal produces a large diminution in the specific electrical conductivity. The diminution produced by the addition of a given quantity of the foreign substance diminishes progressively as further amounts are added. Thus the curve which expresses the conductivity as a function of the percentage composition drops sharply from the value corresponding to either pure metal and has a flat minimum in between. The curve which expresses the thermoelectric power as a function of the composition is entirely similar in character. The reader who is interested in the electrical properties of alloys will find a very good account of the recent developments in *Die elektrischen Erscheinungen in metallischen Leitern* by K. Baedeker (Braunschweig, 1911).

We have seen that a considerable change in electrical conductivity is unaccompanied by corresponding changes in the thermoelectric quantities in the case of pure metals at the melting-point. It seems probable that in this case there is something in the conditions of equilibrium which makes $J+\lambda$ take the same value for the electrons in the two phases. This would not necessarily involve a corresponding equality in the fraction in (48). In addition there is the possibility that liquefaction causes a considerable change, without altering J, in the centres which determine what corresponds to the mean free path.

It is worth while to consider the form taken by (48) in two simple cases. In the first let us suppose that the interior of the conductor may be divided into two classes of regions, for the first of which, τ_1, $\phi\ (= - W)$ has a large value which we can denote by ϕ_0. For the second τ_2 suppose that $\phi = 0$. Then (48) becomes

$$N = \frac{2n}{\sqrt{\pi}} \frac{\sqrt{\frac{\phi_0}{RT}}\,\tau_1 + \frac{\sqrt{\pi}}{2}\,\tau_2}{e^{\phi_0/RT}\tau_1 + \tau_2} = \frac{n}{\sqrt{\pi R}} \left(2\phi_0^{\frac{1}{2}} + \sqrt{\pi} RT \,.\, \frac{\tau_2}{\tau_1}\right) T^{-\frac{1}{2}} e^{-\phi_0/RT} \qquad \text{.........(49).}$$

This formula makes the conductivity increase rapidly with increasing temperature on account of the factor $e^{-\phi_0/RT}$. The experiments made by Horton, Koenigsberger and others on the conductivity of comparatively poor conductors show that the temperature variation can be represented satisfactorily over a considerable range by a formula for the conductivity developed from an expression of type (49) for the number of current-carrying electrons.

In the second place suppose $\dfrac{\phi}{RT}$ to be small. Then if we neglect terms involving $(\phi/RT)^{\frac{5}{2}}$ and higher powers we find

$$N = n\left(1 - \frac{4}{3\sqrt{\pi}} (RT)^{-\frac{3}{2}} \phi_1^{\frac{3}{2}}\right) \qquad \text{..............(50),}$$

where $\phi_1^{\frac{3}{2}}$ is the mean value of $\phi^{\frac{3}{2}}$ taken throughout the volume of the conductor. We should expect this type of formula to apply, qualitatively at least, to the best conductors; for the variation of conductivity with temperature in such cases is comparatively small. According to (50) N will always be less than n, which it approaches asymptotically as ϕ/RT approaches zero. Thus N will always increase with rising temperature. In order to explain the decrease in the conductivity of the pure metals with rising temperature it is necessary, if this theory is to hold, to fall back on an increase either in the number of the centres of force per unit volume or in their strengths.

Some of the rather bad conductors investigated by Koenigsberger exhibit very interesting phenomena. Thus in the case of magnetite the conductivity increases rapidly with increasing temperature at low temperatures, to a maximum at about 240° C. After that it falls off as the temperature rises in a manner which resembles the behaviour of the metals. The rapid increase at low temperatures is evidently due to the large increase of N with T when ϕ/RT is large. The subsequent decrease may tentatively be attributed to an increase in the number of the repelling centres or to a change in the magnetic structure of the substance. A comprehensive account of the temperature variation of the electrical properties of the comparatively poor conductors, by Koenigsberger, will be found in the *Jahrbuch der Radioaktivität und Electronik*, vol. IV. p. 158 (1907).

The Reflexion of Electrons at the Surface of Conductors.

According to the theory on p. 444 *et seq.*, the concentration ν_0 of the electrons at a point close to a conductor in an enclosure at a constant temperature T is determined by T and the intrinsic potential V_m of the conductor. Thus the equilibrium pressure p has a definite value and in accordance with equation (14) is given by

$$p = \nu_0 RT = ATe^{\int \frac{\phi}{RT^2} dT} \quad \text{.................}(51),$$

where ϕ is the internal latent heat of evaporation reckoned per electron and A is a constant characteristic of the material but independent of T. This equation can be established in a very general manner and is true even if the electrons are emitted wholly or in part by the photoelectric action of the complete aetherial (black body) radiation*. Now in accordance with (28) the number of electrons which reach unit area of the surface of the conductor from outside in unit time is

$$N = \nu_0 \left(\frac{R}{2\pi m}\right)^{\frac{1}{2}} T^{\frac{1}{2}} = A\left(\frac{T}{2\pi m R}\right)^{\frac{1}{2}} e^{\int \frac{\phi}{RT^2} dT} \quad \text{......}(52).$$

In the steady state this quantity must be equal to the number emitted by unit area of the substance in unit time provided all the incident electrons are absorbed by the conductor. However, the experiments of von Baeyer† and the writer‡ have shown that a very considerable proportion even of the very slow moving electrons which are emitted by photoelectric and thermionic action is reflected by metals, so that it is necessary for us to take reflexion into account.

---

* See O. W. Richardson, *Phil. Mag.* vol. XXIII. p. 619 (1912).

† *Ber. der Deutsch. Physik. Ges.* Jahrg. 10, p. 96 (1908).

‡ *Phil. Mag.* vol. XVI. p. 898 (1908), vol. XVIII. p. 694 (1909); *Phys. Rev.* vol. XXIX. p. 557 (1909).

If the body is not a perfect absorber then some of the N electrons will be reflected by it. Let the proportion reflected be r. Then the number actually absorbed by the conductor in time dt is $(1-r)Ndt$. Let the number emitted by the same surface in time dt be $\epsilon N dt$. Then the equilibrium condition gives

$$\epsilon = 1 - r \qquad \text{.........................(53).}$$

Thus the emissivity (compared with a perfect absorber) and the reflecting power are complementary.

This result is somewhat analogous to Stewart and Kirchhoff's radiation law. By introducing kinetic theory considerations it can be shown that the equality holds for each group of velocities u, v, w, du, dv, dw[*]. A number of measurements of the reflexion of slow electrons by different metals have recently been made by A. Gehrts[†].

Photoelectric Action.

As is well known, the fact, discovered by Hertz, that a spark passes more easily between two terminals when that which is negatively charged is illuminated by ultra-violet light, led to the discovery that the incidence of light of sufficiently high frequency caused the emission of negative electrons from conductors. This phenomenon, which is called the photoelectric effect, is certainly very general and appears to be a universal property of matter. There is no doubt about the universality of this effect when the term light is understood to include X rays and γ rays, although some experimenters have recently called in question the permanence of the effects exhibited by metals and ordinary ultra-violet light[‡]. We shall now consider what conclusions may be drawn as to the nature of photoelectric action, by the application of the principles of thermodynamics and the kinetic theory of matter.

---

[*] O. W. Richardson, *Phil. Mag.* vol. XXIII. p. 606 (1912).

[†] *Ann. der Physik*, vol. XXXVI. p. 995 (1911).

[‡] G. Wiedmann and W. Hallwachs, *Ver. d. D. Physik. Ges.* vol. XVI. p. 107 (1914). H. Küstner, *Phys. Zeits.* vol. XV. p. 68 (1914).

Consider an enclosure maintained at the constant temperature T containing a material which is photoelectrically active but which has negligible thermionic emission. No such material may exist; but this will not vitiate the results if thermionic and photoelectric actions are independent, a hypothesis which we shall adopt. The equilibrium concentration of the electrons near the photoelectric material will be determined by equation (51), as we may show by considering the work done against a piston which is transparent to radiation but impervious to electrons. The number which return from the enclosure per unit area of the surface of the material in unit time is therefore given by (52). If α is the proportion of these which is absorbed, i.e. not reflected, it is necessary, in order that the conditions should be steady, that the number emitted in unit time should be equal to

$$\alpha A \left(\frac{T}{2\pi m R}\right)^{\frac{1}{2}} e^{\int \frac{\phi}{RT^2} dT} \quad \text{..................(54).}$$

But the number emitted is a function of the intensity or density of the surrounding radiation. Experiments have shown that for monochromatic radiation the number is almost, if not exactly, proportional to the intensity of the illumination and varies little if at all with the temperature, when the other conditions are constant. We may therefore assume the photoelectric emission to be proportional either to the density of the aetherial radiation present or to its rate of emission or absorption. We shall take the latter as being the more general. Our results can then be adapted to the former hypothesis by making the emissivity ϵ equal to unity. If the steady energy density in the vibrations between ν and $\nu + d\nu$ is $L(\nu)\,d\nu$ the energy belonging to these frequencies which is incident on area dS in unit time is

$$\int \frac{c}{4\pi} L(\nu)\, d\nu\, d\omega\, dS \cos\theta,$$

where $d\omega = 2\pi \sin\theta\, d\theta$ and the limits are from 0 to $\pi/2$. Thus, if ϵ is the emissivity of the material, the amount of energy of these frequencies which is emitted from unit area in unit time is $\frac{1}{4} c\epsilon L(\nu)\, d\nu$. Let us suppose that the emission or absorption of unit quantity of radiant energy of frequency ν at temperature T

causes the liberation of $F(\nu, T)$ electrons. Then the total number of electrons emitted in unit time by the complete radiation is

$$N = \frac{c}{4}\int_0^\infty \epsilon F(\nu, T)\, L(\nu)\, d\nu \qquad \text{.................(55).}$$

For equilibrium (54) and (55) must be identically equal for all values of T. Thus if we can express $L(\nu)$ as a function of ν and T, and ϕ as a function of T, we shall have an integral equation of which $\epsilon F(\nu, T)$ is a solution.

We see from equation (26) that ϕ is approximately equal to $\phi_0 + \frac{3}{2}RT$, where ϕ_0 is independent of T. Over the part of the spectrum we are dealing with, $L(\nu)\, d\nu$ can be represented with great accuracy by equation (47) of Chap. XV, or

$$L(\nu)\, d\nu = 8\pi h \frac{\nu^3}{c^3} e^{-\frac{h\nu}{RT}} d\nu.$$

Thus from (54) and (55)

$$\int_0^\infty \epsilon F(\nu, T)\, h\nu^3 e^{-\frac{h\nu}{RT}} d\nu = A_1 T^2 e^{-\frac{\phi_0}{RT}} \qquad \text{.........(56),}$$

where

$$A_1 = \frac{\alpha A c^2}{(2\pi e)^{\frac{3}{2}} m^{\frac{1}{2}} R^{\frac{1}{2}}} \qquad \text{.....................(57).}$$

We shall assume that A_1 is independent of T. This has not been proved, strictly speaking, as ϕ is only approximately equal to $\phi_0 + \frac{3}{2}RT$ and α will also involve T. All that we shall claim for our results, therefore, is that they will represent the way $\epsilon F(\nu, T)$ varies with the parameter ϕ_0 when the matter is supposed to satisfy certain ideal conditions which may be only approximately realized in practice. However, solutions could be found for other cases in which T^2 on the right of (56) is replaced by certain other functions of T.

When A_1 is constant, (56) is solved by

$$\epsilon F(\nu, T) = 0 \quad \text{when} \quad 0 < h\nu < \phi_0 \qquad \text{....................(58),}$$

and

$$\epsilon F(\nu, T) = \frac{A_1 h}{R^2 \nu^2}\left(1 - \frac{\phi_0}{h\nu}\right) \quad \text{when} \quad \phi_0 < h\nu < \infty \quad \text{...(59).}$$

This solution makes the photoelectric action of monochromatic light independent of T, and the emission zero when the frequency is below a certain critical value. These results are both in agreement with experiment.

Now consider the average kinetic energy E_ν of the electrons which are emitted by light of frequency ν. The total energy E emitted under the influence of the complete radiation at T is clearly

$$E = \frac{c}{4}\int_0^\infty E_\nu \epsilon F(\nu)\, L(\nu, T)\, d\nu \quad\text{..............(60).}$$

This must be balanced by the kinetic energy returned to the metal by the stream of electrons, N per unit time, from outside. If n is the number of electrons per unit volume of the space outside, the stream of energy which reaches unit area in unit time is

$$E' = \frac{n}{2}\left(\frac{km}{\pi}\right)^{\frac{3}{2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty um\,(u^2+v^2+w^2)\, e^{-km\,(u^2+v^2+w^2)}\, du\, dv\, dw,$$

and since $k = (2RT)^{-1}$ and $N = n\sqrt{\dfrac{RT}{2m\pi}}$,

$$E' = 2NRT \quad\text{.......................(61).}$$

Of this *energy* let the proportion $1-\beta$ be reflected; then β is the proportion which is absorbed. Thus for equilibrium

$$E = \beta \,.\, 2NRT \quad\text{.......................(62).}$$

If we neglect the effects of reflexion, by putting $\alpha = \beta = 1$, and substitute the value of N given by (47), Chap. XV, making the same approximations as before, we obtain from (60) and (62),

$$\int_0^\infty d\nu\, E_\nu \epsilon F(\nu)\, h\nu^3 e^{-h\nu/RT} = 2A_1 RT^3 e^{-\phi_0/RT} \quad\text{......(63).}$$

Subject to

$$\epsilon F(\nu) = 0 \quad\text{from}\quad \nu = 0 \quad\text{to}\quad \nu = \phi_0/h,$$

and $$\epsilon F(\nu) = \frac{A_1 h}{R^2\nu^2}(1 - \phi_0/h\nu) \quad\text{from}\quad \nu = \phi_0/h \quad\text{to}\quad \nu = \infty,$$

the solution of this is

$$E_\nu = h\nu - \phi_0 \quad\text{when}\quad \phi_0 < h\nu < \infty \quad\text{...........(64).}$$

For values of $h\nu$ which lie between 0 and ϕ_0, E_ν has no meaning, as the corresponding electrons have no external existence.

The solution (64) is dependent on the assumption that the reflexion of electrons can be neglected. This is equivalent to assuming that effects which arise from the collisions of the electrons may be disregarded. But if we consider this from the point of view of the electrons emitted under the influence of the light, we see that neglecting the effect of collisions is tantamount to assuming that the only energy lost by the electrons is used up in overcoming the work ϕ of the forces which tend to retain them in the interior of the substance. Under these circumstances the kinetic energy of the escaped electrons will be equal to that which they acquire by the action of the light (not necessarily from the light directly) minus the work ϕ which they have to do to escape. It is clear that the energy which they acquire under the influence of the light is $h\nu$, where h is Planck's constant. It evidently has the same value for all the electrons liberated by light of the same frequency; any difference in the energy of the electrons emitted by monochromatic light must therefore be attributed to the effect of collisions of the escaping electrons in the interior of the substance.

We can take account of the reflexion of electrons tentatively by putting $\beta = s\alpha$. Then instead of (64) we get

$$E_\nu = s\,(h\nu - \phi_0) \text{ when } \phi_0 < h\nu < \infty \quad \text{.........(65).}$$

For small velocities of incident electrons, such as those with which we are concerned, the proportion reflected increases with increasing energy. It follows that out of a mixed aggregate of incident electrons a greater proportion of the slow ones will be absorbed than of the fast ones, and that the proportion of incident electrons which is absorbed will be greater than the proportion of their incident energy which is absorbed. Thus s will be a positive quantity which is less than unity.

There is no conclusive reason for denying the applicability of the type of argument given above to the emission of material particles of all kinds, whether charged or uncharged and of whatever chemical nature, under the influence of every type of aetherial radiation, provided the chemical actions are of a reversible character so that an equilibrium in the material part of the system can occur.

The foregoing treatment of this subject is taken from papers by the writer*. An equation resembling (64) was first given by Einstein† as a consequence of the view that the energy of light waves was distributed in discrete quanta. For further developments of the subject the reader may consult the following papers:—Einstein, *Ann. der Physik*, vol. XXXVII. p. 832 (1912); *Journal de Physique*, 1913; Planck, *Sitzungsber. d. k. Preuss. Akad. d. Wiss., Math. Phys. Kl.* vol. XVIII. p. 350 (1913); O. W. Richardson, *Phil. Mag.* vol. XXVII. p. 476 (1914).

---

* *Phys. Rev.* vol. xxxiv. p. 146 (1912); *Phil. Mag.* vol. xxiv. p. 570 (1912).

† *Ann. d. Physik*, vol. xvii. p. 145 (1905).

Measurements of the kinetic energy of the electrons emitted by various metals under the influence of light of different frequencies, which have recently been made by the writer and Dr K. T. Compton*, afford considerable support to the above theory. Denoting the maximum observed value of E_ν by E_m, and the mean value by $\overline{E_\nu}$, these quantities were found to satisfy the relations

$$E_m = k_m (\nu - \nu_0)$$

and

$$\overline{E_\nu} = k_\nu (\nu - \nu_0)$$

corresponding to (64) and (65). The values of the constants found are given in the accompanying table:

Metal	Values from E_m			Values from $\overline{E_\nu}$		
	ν_0	λ_0	k_m	ν_0	λ_0	k_ν
Na	51·5	58·3	5·2	52	57·7	2·6
Al	63	47·7	4·3	73	41·1	2·6
Mg	78·5	38·2	5·2	80	37·5	2·55
Zn	80	37·6	5·1	84	35·7	2·8
Sn	83	36·2	4·9	89	33·7	2·75
Bi	91	33	3·55	89	33·7	1·9
Cu	100	30	3·8	97	30·9	1·65
Pt	104	28·8	5·85	103	29·1	2·8

The units are: for ν_0, 10^{13} sec.$^{-1}$: for λ_0, 10^{-6} cm.: and for k_m and k_ν, 10^{-27} erg sec.

According to these results s in (65) is very close to $\frac{1}{2}$ for all the metals investigated. The values of k_m are all somewhat less than the radiation value $h = 6{\cdot}55 \times 10^{-27}$ erg sec., but there are a number of minor causes which might give rise to this discrepancy. There is another way in which h may be estimated from these observations. ν_0, the least frequency which will give rise to any photoelectric emission, is equal to ϕ_0/h. If we may assume that ϕ_0 has the same value as in the thermionic emission and that it is correctly given by the measurements of the latter effect which have been made, we can evaluate h from ϕ_0 and ν_0. A number of observers have found values of ϕ_0 for the thermionic emission from platinum in the neighbourhood of $8{\cdot}32 \times 10^{-12}$ erg. Whence

---

* *Phil. Mag.* vol. XXIV. p. 575 (1912).

$$h = \phi_0/\nu_0 = 8{\cdot}07 \times 10^{-27} \text{ erg sec.}$$

Thus this method of evaluating h leads to numbers which are about as much above the radiation value as the others are below it. From the experimental values of ν_0 the corresponding values of ϕ_0 can be calculated. The differences of ϕ_0 for different metals are found to agree fairly closely with e times the corresponding contact differences of potential. This is in agreement with the theory of contact electromotive force which has already been given.

The linear relation between E_m and ν has been verified independently by A. Ll. Hughes* whose values of k_m exhibit a fair concordance with most of those given in the preceding table. The first measurements of the energy of the electrons emitted under the influence of ultra-violet light were made by Lenard†.

These results must be considered as confirmatory of the validity of Planck's radiation formula.

* *Phil. Trans.* A, vol. CCXII. p. 205 (1912).

† *Ann. der Physik*, vol. VIII. p. 169 (1902).

CHAPTER XIX

TYPES OF RADIATION

In recent years the number of different kinds of radiation with which we have become familiar has been greatly extended. This is especially the case if the expression radiation is made to include any invisible form of energy which originates at a material source and is capable of travelling through empty space at a very high speed. We shall adopt this extension of the term for the present, as it is convenient for the immediate purpose of our discussion.

To classify the different radiations it is desirable to place them in the following groups:

(A) Material and electrically charged.
(B) Material and uncharged.
(C) Aetherial.

To determine whether a given radiation falls in group (A) or not is usually a comparatively simple matter. The rays are bundles of electrified particles in rapid motion. They will therefore behave like an electric current flowing in a flexible conductor and thus be deflected in a magnetic field. The direction of the deflexion enables the sign of the charge which the particles carry to be determined. Another method is to receive a beam of the radiation in an electrically shielded and insulated conducting cylinder. The cylinder will then charge up with electricity of the same sign as that carried by the particles which constitute the radiation. This method, though more direct, is often less easy of practical application than the method which makes use of the magnetic deflexion.

To distinguish between groups (B) and (C) may be extremely difficult. In the case of radiations whose frequencies lie within or close to the range covered by the visible spectrum the occurrence of interference, diffraction, refraction and dispersion is taken to indicate that they are aetherial; since it is only on the theory of waves propagated in a continuous medium that the phenomena enumerated have received an adequate explanation. On the other hand if the frequency of the waves were very great it might be extremely difficult, if not impossible, to detect these effects; so that this criterion would not necessarily be available.

Rather recently W. H. Bragg* has suggested a different method of distinguishing between aetherial and uncharged material rays. The distinction depends upon the geometrical distribution of the *secondary* rays which arise when the *primary* rays fall on matter. Recent investigations have shown that these secondary rays may differ greatly in character according to circumstances. Generally speaking, their nature is determined partly by that of the primary radiation and partly by that of the matter on which it impinges. It often happens that the impact of a simple primary radiation on a chemically simple substance will cause the simultaneous emission of more than one type of secondary radiation. The different types of radiation which may thus arise will be considered more fully below. For the present it is only necessary to realize that the secondary rays in general consist partly of negatively charged particles (electronic or β type) and partly of rays which are similar in their properties to Roentgen rays (X type).

Now consider the emission of β secondary rays which occurs when a beam of X primary rays impinges normally on a slab of absorbing matter of indefinite thinness. The number of β rays emitted from the side of the slab on which the primary rays are incident would be expected to be smaller than the number emitted on the emergent side, if the X rays are uncharged material particles. For, in this case, the emitted electrons will be either those which were originally present in the slab and which are knocked out of it by the moving uncharged particles; or they will consist of those of the moving uncharged material particles which have lost a positively charged constituent by collision with the atoms of the slab. In either event the average value of the component of velocity of the emitted electrons along the normal to the surface will not be zero and will be in the direction of the incident radiation.

---

* *Nature*, vol. LXXVII. p. 270, Jan. 23 (1908); *Phil. Mag.* vol. XVI. p. 918 (1908).

There will therefore be a greater number of β rays emitted from the emergent than from the incident side of the slab.

At first sight it appears that this difference ought not to be found if the incident radiation consists of aetherial pulses similar to those which, as we have seen in Chap. XII, are emitted when a charged body is accelerated. Under these circumstances the natural view of the emission would seem to be that the electrons present in the slab receive an impulse from the electric intensity in the passing pulse. Now the electric intensity is always normal to the direction of propagation of the pulse and thus lies in the plane of the slab. From this point of view there is nothing to favour one side of the slab rather than the other, so that the emission of electrons should be the same on the incident as on the emergent side.

The distribution of the secondary β emission produced by the γ and Roentgen rays has been examined experimentally by Bragg and others. In every case it has been found that there is a larger emission from the emergent side of a thin plate than from the incident side. These results have led Bragg to maintain that the γ and Roentgen rays consist of uncharged material particles and are not "aetherial" pulses. It was shown first of all by O. Stuhlmann*, and about the same time, independently, by R. D. Kleeman†, that a similar lack of symmetry occurs in the emission of electrons from thin plates when illuminated by the ultra-violet light from the arc. It thus appears that Bragg's criterion leads to the view that light, as well as the radiations previously enumerated, consists of neutral material particles.

Under the circumstances it is desirable to reconsider the position. One of the chief difficulties lies in the fact that we have no adequate theory of the mechanism of the absorption of light and similar radiations leading to the emission of electrons. The view which imagines the kinetic energy of the electrons to be derived from the work done by the electric intensity in the pulse, as it passes over them, leads to values of the kinetic energy which are far smaller than those observed experimentally. Thus the occurrence of asymmetrical emission is not the only difficulty experienced by this simple theory. In this connection it is worth while to point out that the simple theory does lead to a slight excess of emergent emission owing to the deflexion of the moving electrons caused by the magnetic force in the aetherial pulse. This deflexion, however, is far too small to produce the observed effects if the force in the pulse is supposed to be distributed uniformly, in the usual way. On the other hand the application of thermodynamic and statistical principles to the study of thermal radiation and photoelectric action led us to the view that when radiant energy causes the disruption of an electron from a material system, the electron acquires an amount $h\nu$ of energy, where $h = 6{\cdot}55 \times 10^{-27}$ erg sec. and ν is the frequency of the radiation. This value is in good accord with experiment, although for any moderate value of ν the amount of energy $h\nu$ is much greater than that which we should expect the electron to acquire, on the simple view discussed above, from the direct action of the pulses. It seems fairly clear either that the (electromagnetic) constitution which we have assumed for the radiation is at fault or the mechanism of the process of absorption is different from what we have supposed.

---

* *Nature*, May 12 (1910); *Phil. Mag.* vol. xx. p. 331 (1910), vol. XXII. p. 854 (1911).

† *Nature*, May 19 (1910); *Roy. Soc. Proc.* A, vol. LXXXIV. (1910).

In view of the latter possibility it is very desirable to see if we cannot find out anything about the magnitude of the Bragg and Stuhlmann effects to be expected from aetherial radiations, without making any definite assumption about the way in which the radiation is absorbed, but keeping to the value $h\nu$ of the energy acquired by the disrupted electrons, which, as we have seen, is confirmed by experiments on photoelectric action and by the theory of heat radiation. Consider again the case of aetherial radiation incident normally on a thin slab of absorbing material. In general, absorption may occur through the operation of processes of very different nature, for example, conduction as opposed to accumulation by relatively fixed and stable arrangements; but we shall suppose that the only type of absorption which we need to consider in our slab is that which results finally in the disruption of electrons from material systems. We shall fix our attention on the state of things which exists after the slab has been illuminated for a sufficiently long time so that there is no further accumulation, in the slab, of energy abstracted from the incident beam. Under these circumstances the energy absorbed from the incident radiation will appear, at any rate in the first instance, as the kinetic energy $h\nu$ which the disrupted electrons possess at the instant of disruption. Thus if there are N of them disrupted in unit time the energy absorbed from the radiation is $Nh\nu$. It is important to observe that the energy of the electrons which we have to consider here is that which they possess before, not after, they are emitted from the slab.

But energy is not the only physical quantity of which the incident beam suffers depletion. We have seen in Chap. X, p. 211, that when a material system similar to that under consideration absorbs an amount E of aetherial energy an amount of momentum E/c disappears from the aether. In the present case this momentum must be communicated to the slab and the electrons it contains, in order to satisfy the law of action and reaction. Now according to the electron theory the action of the radiation is on the electrons; so that this momentum is communicated to them in the first instance, and such of it as is ultimately received by the slab of matter only reaches it indirectly through dynamical actions of the nature of collisions. We shall therefore assume that the electrons receive momentum as well as energy from the incident radiation previous to the occurrence of disruption; although we do not know the precise nature of any process which will communicate an appreciable amount either of momentum or energy from a periodic aetherial disturbance to them. Now consider the accumulation, in any small interval of time, of momentum by the electrons in the slab. The increase of momentum of the slab and contained electrons is due to (1) the momentum of electrons which come into the system, (2) the momentum, reckoned negatively, of the disrupted electrons, and (3) the momentum accumulated during the interval by the electrons present in it. Since the state of the electrons instantaneously present in the slab is steady, by hypothesis, it follows that the difference of (1) and (2) is equal to (3). When (1) is zero, (3) is the momentum derived from the radiation. It follows that the momentum which is acquired by all the absorbing electrons from the radiation, is exactly equal to the momentum of the disrupted electrons at the moment of disruption. But since the energy absorbed is $Nh\nu$ the value of the former amount of momentum is $Nh\nu/c$. If $\bar{u}$ is the average component of velocity of the disrupted electrons, in the direction of incidence of the radiation, an alternative expression for the momentum of the N electrons is $Nm\bar{u}$. Thus

$$\bar{u} = \frac{h\nu}{mc} = \frac{\overline{v^2}}{2c} \quad \text{.........................(1)},$$

if $\overline{v^2}$ is the average value of the square of their velocity at the instant of disruption.

As $(\overline{v^2})^{\frac{1}{2}}$ approaches c, $\bar{u}$ approaches $\frac{1}{2}(\overline{v^2})^{\frac{1}{2}}$, so that for those radiations which give rise to the emission of very high speed β rays there will be a very marked preponderance in the emergent direction. In the case of ultra-violet light, however, for which ν is equal to about 10^{15}, the ratio of $\bar{u}$ to $(\overline{v^2})^{\frac{1}{2}}$ is only about $1:500$. But in this region there is a frequency ν_0 such that if ν is $< \nu_0$ no electrons are able to escape; so that the number emitted is very sensitive to very small changes in the energy communicated to the electrons. Thus a very small difference in the value of the fraction $\bar{u} : (\overline{v^2})^{\frac{1}{2}}$ may make a very considerable difference in the emission of β secondary radiations as between the incident and emergent directions.

It thus appears that the results obtained by Bragg and Stuhlmann may be accounted for without supposing the primary radiations which exhibit them to be of a material nature. There does not therefore appear to be any simple criterion which will invariably enable us to distinguish between aetherial and neutral material radiations. In fact there does not seem to be any convincing evidence of the existence of any radiations which belong to group (B), so that as a working hypothesis it seems most reasonable to classify as aetherial all those radiations which do not belong to group (A). This conclusion is fully substantiated by recent experimental discoveries which have shown that X rays under suitable conditions can exhibit the phenomena of reflexion, diffraction and interference (cf. p. 507).

The nomenclature which is currently employed in describing these radiations is very confusing, as it is not based on any definite system. Thus in different contexts radiations may be differently named, either according to their mode of origin or according to what is believed to be their nature, or according to the effects which they produce. In this way it often happens that radiations, which we have every reason for believing to be identical in nature, have entirely different names depending on more or less fortuitous circumstances. Thus a high speed negative electron is called a β ray if it originates in a radioactive substance, a cathode or Lenard ray if it is produced in a vacuum tube, a secondary Roentgen ray if it is produced by the impact of the Roentgen rays on a solid obstacle, and so on. As it is not our purpose to describe the properties of all the different kinds of radiation in detail we shall extend the scope of the term β ray so as to cover any negatively charged material ray. In the same way we should call a positively charged material ray an α ray, although this term also is usually applied only to those rays which originate with radioactive substances. The uncharged, and presumably aetherial, radiations will be referred to as X rays. We shall use the term γ rays for these when they originate from radioactive substances and Roentgen rays when they are produced in vacuum tubes. For a detailed account of the properties of these radiations the reader may be referred to the following authorities: J. J. Thomson, *Conduction of Electricity through Gases*, Chaps. XI, XII, XIX and XX; Rutherford, *Radioactivity, passim.*

The difference between α and β rays is not merely one of sign. In all cases the specific charge (e/m) has been found to be of a different order of magnitude. For the β rays e/m always has the large value which corresponds to electrons, whereas for α rays e/m is always of the same order as the corresponding quantity in electrolysis. As all the evidence points to the charge e being either equal to, or a small multiple of, the elementary electronic charge in all these cases, it follows that the α rays consist of atoms or molecules which have lost one or more negative electrons. Their properties do not furnish any evidence that we have succeeded in isolating any fundamental positively charged electrical atom which would correspond to the negative electron, unless it be the hydrogen atom which has lost a negative electron.

The most convenient test for the presence of the various types of radiation under consideration is that furnished by the production of electrical decomposition (ionization). Thus the passage of the rays through an insulating gas imparts to it the property of electrical conductivity. The amount of this ionization is often taken to be a quantitative measure of the energy absorbed from the rays. In the case of the X radiations this has proved to be the only method of estimating the energy which is available except in rare cases, although with the α and β radiations more direct methods can usually be employed. The other tests for radiations, in addition to the production of ionization, such as the excitation of fluorescence, photo-chemical action, etc., are probably secondary and depend on electrical decomposition, or at any rate disruption, in the first instance.

The Aether Pulse Theory of the Roentgen Rays.

The suggestion that the Roentgen rays are pulses in the aether, which are produced when the cathode rays are stopped by the walls of the tube or other material obstacle, was first made by E. Wiechert* and Sir G. G. Stokes†. A view, which is really equivalent to this, that these rays are transverse aetherial vibrations of exceedingly short wave-length, had previously been put forward by Schuster‡ and others. The consequences of the aether pulse theory have been worked out by various physicists, including J. J. Thomson§, Abraham‖ and Sommerfeld¶.

We have already seen in Chap. XII that when an electrically charged body or electron is accelerated or retarded, an electromagnetic pulse spreads out in all directions with the velocity of light. In accordance with equation (11) of Chap. XII the total amount of energy in the whole pulse remains constant as it spreads outwards, so that the amount falling on unit area varies inversely as the square of the distance from the source. We also saw that at any point of the pulse the electric and magnetic intensities are equal and mutually perpendicular. They are also at right angles to the direction of propagation. The thickness of a pulse produced by stopping a particle which moves with a given velocity is greater the smaller the acceleration, whilst, at the same time, the energy present in the pulse is less. On this view the main difference between the X rays and thermal radiation or white light is due to the fact that the X rays originate from particles which move with much greater speeds. These high speeds enable the electrons to penetrate right into the interior of the atoms, where they are subject to fields of force much more intense than those to which the slow moving electrons are exposed. Thus in the X ray pulses the forces are both more intense and also change more sharply than those in the light waves. Since, in both cases, the aetherial disturbances are quite irregular, the X rays, on this view, evidently correspond to light waves of high average frequency.

---

* *Ann. der Phys.* vol. LIX. p. 321 (1896).

† *Nature*, p. 427, Sept. 3 (1896); *Mem. Manchester Lit. and Phil. Soc.* vol. XLI. (1896).

‡ *Nature*, p. 268, Jan. 23 (1896).

§ *Phil. Mag.* vol. XLV. (1898).

‖ *Theorie der Elektrizität*, II. Chap. II., Leipzig (1905).

¶ *Sitzungsber. der Kön. Bayerisch. Akad., Math. Phys. Kl.*, Jahrg. 1911, p. 1.

The salient differences between X rays and light are at once accounted for by these considerations. The absence of diffraction and interference under ordinary circumstances is due to the extreme shortness of the equivalent wave-lengths. The absence of refraction is probably to be attributed to the very great difference between the equivalent frequency of the rays and the natural frequency of the bodies traversed; although the fact that the equivalent wave-length appears to be comparable with atomic dimensions may also be an important factor. Either of these causes would probably be competent to explain the absence of refraction, so that there seems to be no difficulty here (cf., however, p. 507).

There is one peculiarity of the aether pulse theory which deserves further consideration. By means of Poynting's theorem and the expressions for the electric and magnetic vectors found in Chap. XII we can write down the density of the stream of energy in different directions from the accelerated electron. In the simplest case*, when the acceleration Γ is in the same straight line as the velocity $v\,(=\beta c)$, the energy radiated across unit area at a point distant r from the accelerated particle, where r makes an angle λ with the direction of v, is

$$E' = \frac{e^2\Gamma^2}{16\pi^2 c^3 r^2} \frac{\sin^2\lambda}{(1-\beta\cos\lambda)^6} \quad \text{..............(2).}$$

In this expression Γ and β are the values of those quantities at the instant when the radiation left the accelerated particle, and therefore at a time $t = r/c$ previous to that at which it reaches the point under consideration. We see from (2) that $E' = 0$ when $\lambda = 0$ or π, so that there is no X radiation along the axis of motion. If β is small the maximum of E' is when $\lambda = \pi/2$, so that the maximum radiation is in the equatorial plane. This is no longer the case when β becomes comparable with unity, i.e. when the velocity of the electron becomes comparable with that of light. In fact as β approaches 1 the value of λ for which E' is a maximum approaches zero, so that the resulting X rays all lie in an infinitely narrow hollow cone about the original axis of motion. Some recent experiments by J. A. Gray* on the X (or γ) rays which are produced when β rays are stopped by matter, show in a convincing manner that the intensity of the resulting X radiation is much greater in the forward than in the backward direction of the β rays. Similar phenomena in the production of X rays by the stoppage of cathode rays had previously been observed by Kaye†. Gray concludes that the asymmetry which he has observed is too large to be accounted for by the aether pulse theory of the production of X rays.

---

* Cf. Sommerfeld, *loc. cit.*

The Scattering of X Rays.

A simple theory of the scattering of X rays by matter has been given by Sir J. J. Thomson‡. When matter is traversed by an aether pulse, the electrons in the former will be accelerated by the electric intensity in the latter. Consider the case in which matter is traversed by a plane pulse in which the electric intensity is imagined to be parallel to the axis of x. Let the direction of propagation of this "primary" pulse be along the axis of z, and let X denote the value of the electric intensity in it. When the primary pulse passes over an electron the latter will be subject to an acceleration $\Gamma = \frac{Xe}{m}$. If we suppose the initial velocity of the electron to be negligible compared with that communicated to it by the pulse, we can put $\sin \epsilon = \epsilon = 0$ in equation (11) of Chap. XII. Thus the energy radiated by the electron in time dt whilst the primary pulse is passing over it is

$$dW = \frac{e^4 X^2}{6\pi m^2 c^3} \frac{dt}{(1-\beta^2)^3} \quad \text{..................(3).}$$

This result will only be true provided Xe is large compared with the forces called into play by the resulting displacement of the electron from its original position of equilibrium. The limitations thus introduced will be considered below. The energy radiated by the electron during the complete passage of the pulse is evidently

---

* *Roy. Soc. Proc.* vol. LXXXVI. p. 513 (1912).

† *Camb. Phil. Proc.* vol. XV. p. 269 (1909).

‡ *Conduction of Electricity through Gases*, Second Edition, p. 321.

$$\int dW = \frac{e^4}{6\pi m^2 c^3} \int \frac{X^2 dt}{(1-\beta^2)^3} \quad \text{..................(4)},$$

where the integrals are extended over the time of passage. It is probable that, except in the case of very penetrating radiations, we shall not be led into serious error if we neglect β^2 compared with unity. Thus

$$\int \frac{X^2 dt}{(1-\beta^2)^3} = E\frac{d}{c} = \frac{W}{c} \quad \text{.....................(5)},$$

where $E = \frac{1}{2}\overline{(X^2 + H^2)} = \overline{X^2}$ is the energy in unit volume of the primary pulse and d is its thickness. If N is the number of electrons in unit volume of the matter, and if the primary pulse only loses energy in this way, the energy $-\delta W$ which is lost by the primary pulse in travelling a distance δz is given by

$$\delta W = \delta(Ed) = -\frac{Ne^4}{6\pi m^2 c^4} W \delta z \quad \text{..............(6)}.$$

Thus the relation between the energy W of unit area of the pulse and the distance z traversed in the matter is

$$W = W_0 e^{-\frac{Ne^4}{6\pi m^2 c^4} z} \quad \text{........................(7)}$$

if W_0 is the value of W when $z = 0$.

On the theory that we are considering, the thickness of a primary pulse is clearly equal to the distance traversed by an electromagnetic disturbance in the time during which the acceleration of the emitting electron is appreciable. This time will be determined by two factors: (1) the velocity of the moving electron relative to the retarding atom, and (2) the geometrical distribution of the field of force, inside the atom, which produces the acceleration. The velocity of the moving electrons is always large and has the value c of the radiation velocity as an upper limit. The linear dimensions of the fields of force will depend upon the constitution of the atom, but will be comparable with the distance between the electrons which are present in the latter. As we shall see later, there are good reasons for believing that the total number of electrons present in the atom is nearly proportional to and comparable with the atomic weight; so that the distance between them will be comparable with, but somewhat smaller than, the diameter of an atom. We should therefore expect that the very penetrating pulses which are produced by the stoppage of very rapidly moving electrons by heavy atoms would have a thickness of the order of 10^{-9} cm. If the stoppage (or acceleration) is produced by light atoms the thickness of the pulses might be ten times as great with electrons moving at the same speed. With slow moving electrons the pulses in each case will be correspondingly thicker, since, other factors being equal, the thickness varies inversely as the speed of the electron. These conclusions would not be materially affected if the pulses were supposed to originate from the acceleration of electrons in the anticathode by the field of force of the moving electrons passing near them, instead of arising, as we have hitherto assumed, from the acceleration of the moving electrons themselves.

When very penetrating X radiations fall on light atoms there is every reason to believe that the pulses are so thin and the electric intensity in them so large that the forces called into play by the resulting displacement of the constitutive electrons of the matter are relatively small. We should therefore expect the theory which leads to equation (7) to hold in such cases. Under the same circumstances the thickness of the secondary pulses will be almost the same as that of the primary pulses. For the thickness of the secondary pulse is equal to the distance travelled by radiation during the time taken by the primary pulse to pass over an electron. But the diameter of an electron is only about 10^{-13} cm. and is therefore practically negligible compared with the thickness even of a very penetrating pulse. It follows that the thicknesses of both the primary and secondary pulses are the same under these conditions and we should therefore expect them to exhibit very similar properties. In confirmation of this result Barkla has found that the "scattered" secondary Roentgen rays, which are produced when penetrating primary Roentgen rays pass over substances made up of light atoms, have the same absorption coefficient as the primary rays, whatever the chemical nature of the matter in which they originate.

For such rays, according to equation (7), the coefficient of absorption is $Ne^4/6\pi m^2c^4$. Since e, m and c are independent of the nature of the matter, the coefficient of absorption for different types of matter made up only of light atoms should be proportional to N, the number of electrons in unit volume. In the case of air Barkla found that the energy, measured by the ionization produced, of the scattered radiation from 1 cubic centimetre is equal to 0·00025 of that of the primary radiation passing through it. Thus, for this substance,

$$Ne^4/6\pi m^2c^4 = 0{\cdot}00025.$$

Since $e/2\sqrt{\pi} = 4{\cdot}81 \times 10^{-10}$, $e/2\sqrt{\pi}mc = 1{\cdot}77 \times 10^7$ and $c = 3 \times 10^{10}$, $Ne = 2\sqrt{\pi}c \times 5{\cdot}95$.

But if n is the number of molecules in 1 c.c. of air, the results of electrolytic experiments show that $ne = 2\sqrt{\pi}c \times 0{\cdot}4327$. Whence $N/n = 14$. Thus the number of electrons divided by the number of molecules is about half the molecular weight. Since the absorption of penetrating X rays by light atoms depends only on the quantity of matter traversed and not on its chemical nature, we are led to the further conclusion that the atomic weights of the different elements are proportional to the number of electrons their atoms contain. Combining the two inferences, it follows that the number of electrons present in the atoms of different elements is a common submultiple of their atomic weights. This conclusion we shall find to be supported by other lines of reasoning (see pp. 490—496).

Polarization of X Rays.

By making use of the properties of the scattered X radiation it is possible to show that Roentgen rays possess features analogous to the polarization exhibited by light and other electromagnetic waves. The existence of this polarization was first demonstrated by Barkla* and his results have since been confirmed by a number of other observers. The principle of these experiments is as follows: Let AB be the direction of the cathode stream in the tube in which the Roentgen rays originate. The cathode rays are stopped by the anticathode at B. Let the Roentgen rays which travel in the direction BC, which is normal to AB, pass through matter consisting of light atoms at C and thus give rise to "scattered" X rays. Now it is probable that the cathode rays are not stopped by a single impact at B so that their motion will become irregular before they cease to cause the emission of Roentgen rays. We should, however, expect the most penetrating rays to arise when the cathode rays are moving at their greatest speed, that is to say, before they have been deflected by many encounters. The main features of the penetrating rays will therefore be determined by the properties of pulses emitted during the acceleration of electrons moving in the direction AB. Let CE be a line parallel to AB in the plane of ABC and let CD be perpendicular to EC, CB and BA. The acceleration at B of an electron whose velocity is along AB will give rise to an element of pulse travelling along BC in which the electric intensity is parallel to CE and the magnetic intensity is parallel to CD. If this pulse falls on an electron at C, the latter will be subject to an acceleration along CE. If we assume the electron to be at rest initially, the secondary pulse will have zero intensity along the axis CE and its maximum intensity in the equatorial plane through C which contains CD and BC. If the cathode rays were stopped by the first encounter at B the secondary rays would have zero intensity along the polar axis CE and a maximum intensity at points in the equatorial plane through C. If the primary tube were rotated so that AB became parallel to CD, the secondary rays would fall to zero in the direction of CD, and CE would become one of the directions of maximum intensity. Since the electrons are believed to be only partially stopped by a single collision the intensity of the secondary rays will never drop to zero. We shall only be able to observe a minimum value in the directions indicated as compared with the maximum value in the perpendicular plane.

---

* *Phil. Trans.* A, vol. CCIV. p. 467 (1905).

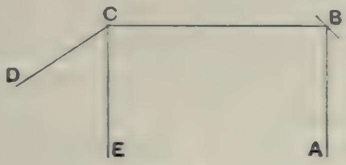


Fig. 51.

The observations alluded to are in satisfactory accordance with the theory just outlined. The maximum difference of intensity in the directions CE and CD observed by Barkla is about 20 %. The occurrence of polarization phenomena is not easy to account for on the material theory of X rays, although such a position has been defended by Bragg*.

The Scattering of α and β Rays by Matter.

When these very rapidly moving particles traverse matter it seems fairly certain that they pass through the interior of the atoms and not simply through the spaces between them. For instance, it is found that when a pencil of α rays is made to pass through sufficiently thin sheets of solid substances only a small fraction of them are either deflected through appreciable angles or stopped, whereas a calculation based on the fairly concordant values of the dimensions of molecules given by various methods shows that the chance of a particle passing through such a sheet without colliding with a molecule is excessively slight. In fact, since Rutherford's experiments have shown that the α particles from radioactive substances are charged atoms of helium, we have every reason to expect that they are of atomic dimensions themselves, and it is probable that they would be geometrically unable to pass through a sheet of solid of the kind contemplated without intersecting the atoms of the latter.

Admitting that the moving charged particles pass through the interior of the atoms, in many cases without sensible deflexion, it follows that such deflexions as occur must arise from encounters with systems forming part of the atom rather than with the atom as a whole. We should therefore expect that a study of the scattering of pencils of these particles which results from their transmission through matter would afford valuable indications toward a knowledge of the structure of the atoms of the latter; although it is possible that the effects to be expected might not be very dissimilar even if widely different views of the structure of the atoms were adopted. The problem has been considered theoretically by both Thomson† and Rutherford‡, who have adopted rather divergent views both as to atomic structure and as to the nature of the deflexions. Let us consider Thomson's treatment first.

* *Phil. Mag.* vol. xiv. p. 429 (1907).

† *Camb. Phil. Proc.* vol. xv. p. 465 (1910).

‡ *Phil. Mag.* vol. xxi. p. 669 (1911).

Thomson considers the atoms through which the charged particles pass to be made up of N_0 negative electrons accompanied by an equal quantity of positive electricity. The deflexion of a moving negative electron, for example, will then arise from two causes: (1) the repulsion of the electrons distributed through the atom, and (2) the attraction of the positive electricity. In regard to (2) two cases are considered: (1) the positive electricity is distributed with uniform volume density throughout a sphere whose volume is equal to that of the atom (see Chap. XXI), and (2) it is divided into separate equal units, as to which it is considered probable that they occupy a much larger volume than that occupied by an electron.

The deflexion of an individual electron due to its passage through a single atom will depend on the manner in which the atom is approached, but the average value of this deflexion taken over a large number of encounters will be a definite quantity θ. If we consider a sufficiently large number of consecutive deflexions it is probable that no serious error will be introduced if they are all treated as equal to the mean value θ. It is to be remembered that the direction of any individual deflexion is a matter of chance and its magnitude is very small. By considering the resolution of any large number n of such deflexions θ along any two perpendicular axes in a plane at right angles to the original motion, it is evident that the problem of finding the average effect of the n deflexions is the same as that of finding the average amplitude of the resultant of n vibrations, the amplitude of each of which is θ and the phase of each of which is entirely fortuitous. This problem has been solved by Lord Rayleigh* who has shown that the average resultant amplitude is

$$\sqrt{n}\theta \quad \text{..............................} (8).$$

It follows that if the rays pass through a plate of thickness t, containing N atoms, of radius b, per unit volume, the average deflexion which they experience is $\theta\sqrt{N\pi b^2 t}$. The problem is thus reduced to the determination of the value of θ.

The part θ_1 of θ which arises from deflexions caused by the negative electrons can be found by means of the theory of particles moving under central forces which vary inversely as the square of

* *Theory of Sound*, Second Edition, vol. I. p. 35.

the distance*. When the velocity V of the β particles is so large that the resulting deflexion is small, the deflexion produced by a single encounter is $\frac{2e^2}{mV^2}\frac{1}{x}$, where x is the perpendicular distance between the electron encountered and the undeflected path. The average value of this for all the electrons which lie within a distance a of the line of motion is $\frac{4e^2}{mV^2}\frac{1}{a}$. If l is the average length of the path of the moving particle which lies inside the atom, if the electrons are uniformly distributed in the atom and if there are ν of them per unit volume of the latter, the number of collisions with electrons which lie within a distance a is $\nu\pi a^2 l$. If we suppose for the moment that the negative electrons are alone operative and consider any very large number σ of encounters with atoms the average total deviation would be

$$\frac{4e^2}{mV^2}\frac{1}{a}\sqrt{\sigma\nu\pi a^2 l} = \frac{4e^2}{mV^2}\sqrt{\sigma\nu\pi l}.$$

Since this must be $\sqrt{\sigma}$ times the effect of a single encounter the effect of the latter is

$$\theta_1 = \frac{4e^2}{mV^2}\sqrt{\nu\pi l} = \frac{16}{5}\frac{e}{mV^2}\frac{1}{b}\sqrt{\frac{3N_0}{2}} \quad \text{...........(9)},$$

substituting $\sqrt{l} = \frac{4}{5}\sqrt{2b}$, where b is the radius of an atom.

If ϕ_1 is the average deflexion due to the sphere of positive electrification the theory of central forces shows that this is given by

$$\phi_1 = \frac{e^2}{mV^2}\frac{N_0}{b}\frac{\pi}{4} \quad \text{.....................(10)},$$

provided ϕ_1 is small. When the positive electricity is made up of definite units the mean deflexion ϕ_2 due to these is given by

$$\phi_2 = \frac{16}{5}\frac{e^2}{mV^2}\frac{1}{b}\sqrt{\frac{3N_0}{2}}\sqrt{1-\left(1-\frac{\pi}{8}\right)\tau^{\frac{1}{3}}} \quad \text{......(11)},$$

where τ is the ratio of the volume occupied by the positive electricity to the volume of the atom.

Thus, according as we adopt the first or second hypothesis as to the geometrical configuration of the positive electricity, the value of θ will be given by the equations

$$\theta = (\theta_1^2 + \phi_1^2)^{\frac{1}{2}} \text{ or } \theta = (\theta_1^2 + \phi_2^2)^{\frac{1}{2}} \quad \text{............(12)},$$

---

* Routh, *Dynamics of a Particle*, Cambridge, 1898, Chap. VI.

and the average deflexion ψ_m in passing through a thin plate of thickness t will be given by either

$$\psi_m = \frac{e^2}{mV^2}\left\{\frac{384}{25}N_0 + \frac{\pi^2}{16}N_0^2\right\}^{\frac{1}{2}}\sqrt{N\pi t} \quad\text{..............(13)},$$

or

$$\psi_m = \frac{e^2}{mV^2}\left\{\frac{384}{25}N_0\left[2-\left(1-\frac{\pi}{8}\right)\tau^{\frac{1}{3}}\right]\right\}^{\frac{1}{2}}\sqrt{N\pi t} \quad\text{...(14)}.$$

As the only unknown quantities on the right-hand sides of (13) and (14) are N_0 and τ, if the equations are supported by experiment they ought to enable us to determine the value of N_0 the number of electrons in the atom.

The extent to which these equations are in accordance with the observed scattering of β rays when they pass through thin sheets of various solids has been examined by Crowther*. If ψ is any particular angle of deflexion it follows from the theory of errors that the probability of a deflexion greater than ψ is $e^{-\psi^2/\psi_m^2}$. Thus from (8) the thickness t_0 for which this probability is $\frac{1}{2}$ will be given by

$$e^{-\psi^2/ct_0\theta^2} = \tfrac{1}{2} \quad\text{......................(15)},$$

or

$$\psi/t_0^{\frac{1}{2}} = \theta\sqrt{(c\log 2)} \quad\text{...................(16)},$$

where c is a constant for any particular substance. The probability that the deflexion is less than ψ is equal to $1-e^{-\psi^2/ct\theta^2}$, that is to

$$1-e^{-k/t} \quad\text{.........................(17)},$$

where k is constant for any particular substance, if ψ is kept constant. From (8) and (16)

$$\psi/t_0^{\frac{1}{2}} = \psi_m/t^{\frac{1}{2}} \times \sqrt{\log 2} \quad\text{................(18)},$$

and from either (13) or (14)

$$\psi/t_0^{\frac{1}{2}} = \text{const.} \times \frac{1}{mV^2} \quad\text{................(19)},$$

where the absorbing medium is kept the same and the velocity of the incident rays is varied. When ψ is also kept constant

$$mV^2/t_0^{\frac{1}{2}} = \text{constant} \quad\text{...................(20)}.$$

The method of investigation adopted by Crowther was to measure the proportion of the incident β rays whose deflexions were less than a fixed value ψ when they passed through absorbing sheets of different materials and when the thickness of the sheets and the velocity of the rays were varied. In this way he has been able to confirm equations (16), (17) and (20) using aluminium sheets, and equation (16) with platinum sheets also. The value of $\psi/t_0^{\frac{1}{2}}$ enables N_0 to be calculated from (13) and (14). In reducing (14) τ is assumed equal to zero, but it is shown that the value assumed for τ cannot affect the results vitally. If (13) is taken, N_0 is found to be proportional to the atomic weight W of the elements C, Al, Cu, Ag and Pt, the value of N_0/W varying only between 2·87 and 3·32. This variation is comparable with the error of measurement. If (14) is taken, N_0/W ranges from 3·7 for carbon to 33·5 for platinum. Since the scattering of X rays by matter leads to the conclusion that the number of electrons per atom is proportional to the atomic weight, the hypothesis underlying (14) is discarded and the experiments are taken to establish the view that the positive electricity is uniformly distributed. Subject to this hypothesis the results show that the number of electrons present in any atom is equal to three times the atomic weight, within the limits of accuracy of the measurements. Although this estimate is considerably higher than that given by the scattering of Roentgen rays by light atoms it is satisfactory to note that it is of the same order of magnitude.

---

* *Roy. Soc. Proc.* A, vol. LXXXIV. p. 226 (1910).

Rutherford's treatment of the scattering differs from Thomson's in two important particulars. He attributes the main features of the scattering of the particles when they pass through thin layers of matter to the effects of single encounters (single scattering) and not to the chance combination of a multitude of excessively minute deflexions. This is the most radical difference. It is supported by the experiments of Geiger* and others on the scattering of α rays produced by thin sheets of matter. These experiments have shown that the proportion of α particles which are scattered through large angles is very much larger than that which would be expected as the cumulative effect of a multitude of small deflexions. The other difference is in the hypothetical constitution of the atom which is adopted. This is imagined to consist of a central point charge $\pm N_0 e$ surrounded by a sphere of uniformly distributed electrification equal altogether to $\mp N_0 e$.

---

* Geiger and Marsden, *Roy. Soc. Proc.* A, vol. LXXXII. p. 495 (1909); Geiger, *Roy. Soc. Proc.* A, vol. LXXXIII. p. 492 (1910).

The assumption of a large central charge is shown to be required to account for the large individual deflexions.

The probability of a single deflexion greater than a given angle is clearly determined, making use of the theory of central forces, by the chance that the undeflected path of a particle, moving with given velocity, will pass within an assigned distance of the fixed centre. In this way it is shown that the fraction ρ of the particles which are deflected through angles lying between ϕ_1 and ϕ_2 is

$$\rho = \frac{\pi}{4} ntb^2 \left(\cot^2 \frac{\phi_1}{2} - \cot^2 \frac{\phi_2}{2} \right) \quad \text{...........(21)},$$

where

$$b = \frac{2N_0 eE}{mV^2} \quad \text{.......................(22)},$$

E being the charge, m the mass and V the velocity of the α particle. The results of this theory are shown to be in satisfactory accordance with

(1) the measurements of the proportion of α particles scattered through large angles;

(2) the scattering of α particles through all angles, if allowance is made for the effects arising from the cumulative effects of small deflexions considered by Thomson;

(3) the proportion of α particles scattered through large angles by sheets of different elements, if the central charge is assumed to be proportional to the atomic weight.

It appears that Rutherford's calculations, as well as Thomson's, lead to the constancy of the fractions $\psi/t_0^{\frac{1}{2}}$ and $mV^2/t_0^{\frac{1}{2}}$, for the scattering of the β rays under the conditions investigated by Crowther. The values of the constants, however, are different in the two cases and so are the estimated values of N_0. Using Crowther's data Rutherford calculates the following values for N_0, the number of electrons which would have a total charge equal to the hypothetical central charge: for Al, $N_0 = 22$; for Cu, $N_0 = 42$; for Ag, $N_0 = 78$; and for Pt, $N_0 = 138$. These values are roughly in the proportion of the atomic weights. Platinum and gold appear to be the only materials for which suitable data relating to the scattering of the α particles through large angles are available; but in that case the resulting estimate of N_0 is in substantial agreement with that deduced from the behaviour of the β rays.

On the theory of single scattering the number of particles deflected through an angle greater than a given angle will be equal to $k't$, where k' is a constant, if the thickness t is small. Thus, instead of (17), the proportion I/I_0 of the rays for which the deflexion is less than an assigned amount will be given by

$$I/I_0 = 1 - k't \quad\text{..........................(23)},$$

provided t is small enough. Since Crowther found (17) to be confirmed by his experiments this test is in favour of Thomson's theory. This divergence might be reconciled by supposing that the cumulative effect of the deflexions of β rays by electrons is relatively more important than the compound deflexion of the α rays. This seems reasonable, as a moving particle will come within range of many more electrons than atoms. The moving electrons will be deviated in varying degree both by the electrons and the large central charge, whereas the α particles will only be affected appreciably by the central charge, on account of the small mass of the electrons.

It is worth while pointing out that Rutherford's estimates of N_0 agree better with those of the number of electrons per atom given by the scattering of Roentgen rays than do the estimates of N_0 deduced from Thomson's theory.

Secondary Rays.

The secondary rays which are emitted when Roentgen rays are absorbed by matter possess many interesting properties. The scattered radiation, which is similar to the primary radiation, has already been considered. In general two additional types of secondary radiation are found to be emitted. One of these is of the X type and is called, for reasons which appear below, "characteristic" radiation; and the other is of the β type.

Let us consider for a moment the means which are available for recognizing and classifying what we may provisionally call pure radiations of the X type. Suppose that we are dealing with primary radiations whose penetrating power is comprised within the limits of that of the rays given off in quantity by an ordinary Roentgen ray bulb. When such rays are allowed to fall on matter it is found that characteristic radiations are not excited, provided the matter consists entirely of elements whose atomic weights are all below 40. If the atomic weights exceed 40, characteristic radiations are found to be emitted. In all cases there is evidence of the occurrence of scattered radiation. These facts would lead us to conclude that absorption of the rays by light elements is a less complex phenomenon than absorption by matter containing the heavier atoms. Let us take as our criterion of a pure radiation the condition that its law of absorption, by matter consisting only of light elements, should be an exponential one. That is to say, if the incident intensity is i_0 the intensity i, after traversing normally a thickness x of the matter, will be

$$i = i_0 e^{-kx} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(24),$$

where k is independent of x. This law results, of course, if successive equal thicknesses of the absorbing matter absorb equal fractions of the radiation entering them. The intensity i is taken to be measured by the ionization which the rays will produce in a thin layer of a gas like air, which is composed solely of elements of low atomic weight. This method is satisfactory because the amount of ionization in such cases has been shown to be proportional to the loss of intensity due to absorption. Now suppose that we have a number of different pure radiations and that we determine the value of k for each of them when they are absorbed by sheets of some particular light element. Let the resulting values of k be λ_1, μ_1, ν_1, etc. Let the corresponding values of k for some other light element be λ_2, μ_2, ν_2, etc. Then it is found that $\lambda_1 : \mu_1 : \nu_1$ etc. $= \lambda_2 : \mu_2 : \nu_2$ etc. No such simple relation between the values of k is found to hold when absorption by materials with heavier atoms is considered, so that this result favours the comparative simplicity of the phenomena in the case of the lighter atoms.

It is clear that if we take any particular element of low atomic weight the value of the absorbability k will give us a means of classifying any particular pure radiation. As it is very readily obtainable in thin sheets, aluminium is generally taken as the standard. This method of investigation has been used by Barkla* and his collaborators, whose researches have greatly increased our knowledge of the X type of secondary Roentgen rays. We shall see below that there is another criterion which may be used instead of k, namely the maximum velocity of the β rays emitted when the X rays fall on metals. This has been shown to have, for each pure radiation, a definite value which is independent of the metal (cf. also p. 510).

---

* Barkla, *Phil. Mag.* vol. XI. p. 812 (1906); Barkla and Sadler, *Phil. Mag.* vol. XVI. p. 550 (1908); vol. XVII. p. 739 (1909); Barkla, *Jahrb. der Radioakt. etc.* vol. III. p. 246; *Phil. Mag.* vol. XXII. p. 396 (1911).

When the primary rays are such as are emitted by an ordinary Roentgen ray bulb the difference between the secondary rays emitted by aluminium (atomic weight 27) and an element like copper (atomic weight 63) is very striking. In the first place, the quantity of the secondary emission (scattered radiation) from aluminium is comparatively insignificant. The amount of the secondary radiation from the copper is very much larger. The two radiations differ also in quality. We have seen that the aluminium radiation has the same penetrating power as the primary radiation. The copper radiation is much more absorbable. It is also very nearly a pure radiation. The current evidence is to the effect that it consists almost entirely of a pure radiation mixed with a trace of scattered radiation. The amount of the scattered radiation from copper, and other elements of atomic weight above 40, appears to be of the same order as that given out by aluminium.

This pure secondary radiation is that which we have referred to as the characteristic secondary radiation. It is characteristic in the sense that its absorbability k has a value which is characteristic of the metal from which it is emitted. In the case of the elements whose atomic weights lie between about 40 and 100 the absorbability of the characteristic secondary rays is quite independent of the nature of the primary X rays which are used to excite them. With the elements of still higher atomic weight the phenomena are more complicated and it appears that these elements give off more than one kind of characteristic radiation. It is convenient to allude to the characteristic radiation from a particular element M as the M-X rays. Thus the characteristic rays from copper are called the Cu-X rays. This notation is due to Bragg.

The absorbability of the characteristic rays diminishes, and their penetrating power increases, as the atomic weight of the parent element increases. The characteristic rays differ from the scattered rays in one important particular in addition to those already mentioned. They show no trace of polarization. This leads to the view that the manner in which the primary rays cause the formation of the characteristic rays is less direct than that in which the scattered rays are produced. This conclusion may be considered to be established when we recall that the scattered rays are similar in their properties to the primary rays and independent of the matter in which they originate: whereas the characteristic rays have their properties determined solely by the matter of origin and not at all by the character of the primary rays.

We shall now turn to the absorption of pure X radiations by sheets of different elements. A series of results obtained by Barkla and Sadler is exhibited in the accompanying diagram

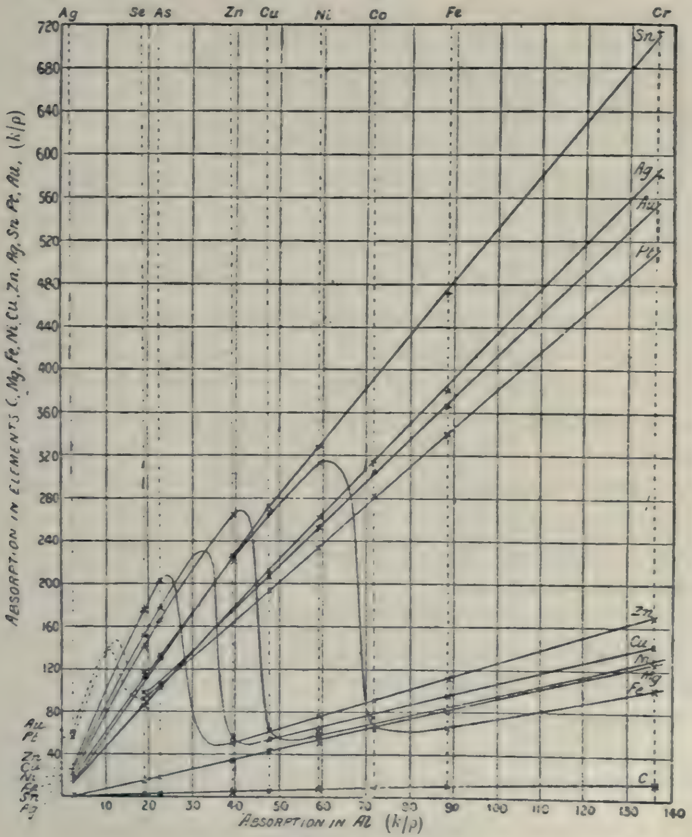


Fig. 52.

(Fig. 52). The abscissae are proportional to the absorption coefficient of the various radiations in aluminium. The experimental values of k/ρ in aluminium are indicated by the abscissae corresponding to the various vertical dotted lines alongside which are written the chemical symbols of the different elements by which the pure characteristic radiations were emitted. The ordinates are the values of k/ρ, where ρ is the density of the absorbing element, for the elements C, Mg, Fe, Ni, Cu, Zn, Ag, Sn, Pt and Au. In the case of the elements of atomic weight below 40, namely C and Mg, the relation between the absorbability of the radiations by the respective elements and that by aluminium is a linear one, as has already been pointed out on p. 497. The same thing, however, will be observed to be true in the case of Pt*, Au*, Ag and Sn as well. In fact such a linear relation between the absorption coefficients for different pure radiations has been found to hold quite generally, provided the range of radiations tested neither includes that characteristic of the absorbing substance nor lies near the less absorbable side of it.

The curves for Fe, Ni, Cu and Zn are quite different. As they all exhibit the same features we need only consider one of them, that for nickel for example. The absorbability by nickel of the radiations from Cr, Fe, Co and Ni, which are either more absorbable than, or as absorbable as, that characteristic of nickel, is proportional to the absorption of the same radiations in aluminium. The next most penetrating radiation, as measured on the aluminium scale, is the Cu-X radiation, and this shows a small but definite increase over the absorbability required by the law of proportionality. As the absorbability in terms of aluminium of the characteristic rays diminishes further, their absorbability in nickel increases very rapidly. Thus the absorption coefficient of the Zn-X rays in nickel is several times that of the Cu-X rays, although the atomic weight of zinc (65) is only 2 units greater than that of copper. Further diminution in the aluminium absorbability of the characteristic rays is accompanied by an approximately proportional diminution in the absorbability of the rays in nickel. Thus the curve again approximates to a straight line which has a different slope from the linear portion on the more absorbable side of the radiation which is characteristic of nickel. It will be noticed that in each case the radiation having the maximum absorbability is that which is characteristic of the element next but one to the absorbing element when they are arranged in the order of increasing atomic weight.

---

* It is necessary to except the penetrating Ag-X and Se-X rays in these cases.

Direct experiments have shown that the abnormal absorption of radiations which are slightly more penetrating than the radiation characteristic of the substance tested, is the cause of the emission of the characteristic radiation. The *amount* of the characteristic radiation emitted by an element is found to be zero until the exciting rays become at least as penetrating as the characteristic radiation; as the penetration, measured in terms of aluminium or some substance which behaves similarly to aluminium, is increased beyond this critical value, the amount of characteristic radiation which the exciting radiation will cause to be emitted, compared with the ionization which it produces in air or any other gas containing only light atoms, increases at first to a maximum value and then diminishes. In fact, the quantity of characteristic radiation emitted appears, for all exciting radiations, to be proportional to the excess of absorption over what would be given by the linear relation which holds for exciting radiations that are more absorbable than the characteristic radiation.

Since these results appear to be quite general, the occurrence of humps, like those shown in the relative absorbability curves for Zn, Cu, Ni and Fe in Fig. 52, affords a simple and delicate method of detecting the presence of unknown characteristic radiations. Working in this way Barkla and Nicol* have succeeded in separating two quite distinct characteristic radiations from each of the elements silver, antimony, iodine and barium. A comprehensive examination of all the radiations thus far discovered shows that, when they are arranged in the order of k/ρ, their comparative absorbability in aluminium, they fall into two series. These have been called by Barkla the K and L series respectively. In each series the order of diminishing absorbability is that of increasing atomic weight of the metal of origin. The data for the radiations which have thus far been isolated are given in the following table:—

---

* *Nature*, Aug. 4 (1910).

Radiation	Atomic Weight	Absorbability of Fluorescent Radiation (k/ρ in Al)		Remarks
		Series K	Series L	
H-Mg	1·008—24·32			No radiation observed. Probably very absorbable.
Al and S	27·1 and 32·07			Radiation observed but k/ρ not yet measured.
Ca	40·09	435		Less accurate than succeeding values.
Cr	52	136		The value in brackets is deduced by Barkla from X ray measurements. The value for Rb has not been determined so accurately as the others.
Fe	55·85	88·5		
Co	58·97	71·6		
Ni	58·68 (61·3)	59·1		
Cu	63·57	47·7		
Zn	65·37	39·4		
As	74·96	22·5		
Se	79·2	18·9		
Br	79·92	16·4		
Rb	85·45	13·7		
Sr	87·62	9·4		
Mo	96	4·7		
Rh	102·9	3·1		
Ag	107·88	2·5	700	Values for L more accurate than for K series.
Sn	119	1·57		
Sb	120·2	1·21	435	
I	126·92	0·92	306	
Ba	137·37	0·8	224	
Ce	140·25	0·6		
W	184		33	These values are comparatively approximate.
Pt	195		27·5	
Au	197·2		25	
Pb	207·1		20	
Bi	208		19	

Reasons have been given by Barkla for believing in the existence of series other than the series K, L which have already been explored*. It will be noticed that the fact that the characteristic radiations are only excited by radiations more penetrating than themselves is analogous to Stokes's law in optics: according to which fluorescent light is invariably of lower frequency than the light which excites it. For this reason the characteristic radiations have sometimes been called fluorescent radiations. It is obvious that each pure characteristic radiation has features which are closely analogous to those exhibited by a monochromatic emission of fluorescent light, and the sum of the radiations characteristic of any given element may be regarded as a fluorescent spectrum in the X ray region.

---

* *Phil. Mag.* vol. XXII. p. 408 (1911).

The Relation between β ray and X ray Emission.

We have seen that X rays are emitted when high-speed electrons impinge on matter and also that when X rays are absorbed by matter secondary radiation of the β type is emitted. Recent investigations have shown that this transformation of X ray into β ray energy and the inverse change obey a number of fairly simple laws which appear to be of a general character.

The specification of the nature of any single β ray is a quite simple matter. Since each ray is a moving electron, its nature is determined when its speed and direction of motion are given. In general a beam of β rays consists of a shower of electrons having different speeds and directions of motion. By the insertion of suitable stops the range of the directions of motion can be restricted to any desired extent. By the application of a magnetic field the β rays can also be spread out into a sort of spectrum. In this way sufficiently homogeneous beams of β rays can be obtained for experimental purposes. The β rays which are given out by radioactive substances or which are emitted when X rays are absorbed by matter are, as a rule, of a heterogeneous character. This heterogeneity arises in part from the fact that the rays originate at different depths in the matter and thus lose varying amounts of energy before they emerge. In many cases it has been shown that the maximum energy of the β rays which thus arise is a definite quantity which exhibits rather simple relationships.

In the cases now under consideration it has not always been found possible to measure the energy of these rays by the comparatively direct methods involving stoppage by an electric or deflexion by a magnetic field, which have been employed in the case of the cathode rays and of the electrons emitted by hot bodies and by bodies illuminated by ultra-violet light. The reason for this is either that the energy of the rays is too great to be much affected by the electric fields at our disposal or else that the rays are not concentrated enough for the magnetic method to be available. A measure of the energy of the β rays has been found in the distance that they are able to travel in a gas like air before they cease to produce additional ionization of the gas. This distance has been found to be sufficiently definite to be taken as a satisfactory index of the velocity of the quickest rays in any given group.

Using this criterion Sadler* and Beatty† have been able to show that the maximum velocity of the β rays emitted when a given characteristic X radiation is allowed to fall on different metals always has the same value. This maximum velocity is therefore characteristic of the X radiation and independent of the metal; it is greater the greater the atomic weight of the metal of which the X rays are characteristic.

This question has been studied more fully by Whiddington‡. He has shown by experiment that if the original velocity of a group of β rays is v_0, their velocity v after traversing a thickness d of matter is given by

$$v_0^4 - v^4 = ad \quad \text{......................(25)},$$

where a is a constant characteristic of the matter. Since (25) appears to hold with fair approximation down to the velocity (approximately $v = 0$) at which ionization ceases, it follows that the velocity v_s of the fastest secondary rays is given by

$$v_s^4 = ad \quad \text{..........................(26)},$$

where $a = 2 \times 10^{40}$, for air at atmospheric pressure, and d is the maximum distance in which ionization is perceptible. Analysing the results of Sadler and Beatty in a manner which depends on this principle it appears that

$$v_s = k'w \quad \text{..........................(27)},$$

where k' is a constant and w is the atomic weight of the radiator of which the secondary rays are characteristic. This relation has been verified for the Fe-, Cu-, Zn-, As-, Sn-, Mo-, and Ag-X rays. The value of k' is very close to $10^8 (\pm 4\,\%)$.

We have seen that the characteristic X rays are only excited by X rays more penetrating than themselves. Using an X ray tube with a silver anticathode Whiddington§ has investigated the relation between the velocity of the cathode rays in the primary tube and the amount of secondary characteristic radiation emitted when the resulting primary X rays are allowed to fall on different metals. He finds that no characteristic secondary rays are emitted until the cathode rays have a critical minimum velocity v_p. The emission begins quite sharply at this point and subsequently varies as the fourth power of the velocity of the cathode rays*. The value of v_p is connected with the atomic weight w of the secondary radiator by the simple relation

---

* *Phil. Mag.* vol. XIX. p. 337 (1910).

† *Phil. Mag.* vol. XX. p. 320 (1910).

‡ *Roy. Soc. Proc.* A, vol. LXXXVI. pp. 360, 370 (1912).

§ *Roy. Soc. Proc.* A, vol. LXXXV. p. 323 (1911).

$$v_p = kw \quad \text{..........................(28).}$$

The value of k is very close to 10^8 and, in fact, $k = k'$ within the limits of error of the experiments.

Thus primary cathode rays of velocity just greater than $v_p = kw$ give rise to primary Roentgen rays, and these if absorbed by a metal of atomic weight w would give rise to secondary Roentgen rays, characteristic of that metal, which give rise to secondary β rays whose maximum velocity is $v_s = k'w = kw = v_p$. The maximum velocity of the secondary β rays is equal to the minimum velocity of the primary cathode or β rays which are able, indirectly, to excite the characteristic Roentgen rays which caused the emission of the secondary β rays.

The formulae above only apply without modification to Barkla's K series of characteristic radiations. According to Whiddington† the case of both the K and L series is covered by the more general formulae

$$v_p = k\,(Aw + B) \quad \text{......................(29),}$$
$$v_s = k'\,(Aw + B) \quad \text{.....................(30).}$$

For the K series $A = 1$ and $B = 0$.

For the L series $A = \frac{1}{2}$ and $B = -25$.

Thus the secondary cathode rays, excited by the characteristic rays which belong to the L series from an element of atomic weight w, have a maximum velocity equal to

$$\left(\frac{w}{2} - 25\right) 10^8 \text{ cm. sec.}^{-1}.$$

It will be observed that there is a very close resemblance between the emission of electrons by matter under X ray and ultra-violet light illumination. In both cases the number of electrons emitted is proportional to the incident intensity, whereas the maximum energy of the electrons is independent of this intensity. These facts receive a simple and obvious explanation on the view that the X rays and light consist of showers of material particles or of bundles of energy which are localized in space and do not spread out as the distance from the source is increased. On the other hand in the case of light (and also X rays, see p. 507) it is difficult to account for the phenomena of interference and refraction on such a view. It seems a little safer, therefore, to suppose tentatively that the energy of the emitted electrons in some way represents a condition which determines the disruption of matter under the stimulus of a given radiation. To agree with the results of the theory of black body radiation and of the experiments on the emission of electrons under the influence of light it is necessary that one part of this condition should be that the energy of the disrupted electrons is either equal to $h\nu$, where $h = 6{\cdot}55 \times 10^{-27}$ erg sec. and ν is the frequency of the radiation, or is an integral multiple of this quantity. It also appears that this condition must be of a very general character and necessarily inherent in all types of matter.

---

* Cf. J. J. Thomson, *Phil. Mag.* vol. xiv. p. 217 (1907).

† *Roy. Soc. Proc.* A, vol. LXXXVI. p. 378 (1912).

We have seen that the maximum energy of the electrons emitted under the influence of illumination by light is given by the equation

$$T_m = \nu h - w_0 \quad \text{.......................(31)},$$

where w_0 is a constant characteristic of the type of matter by which the light is absorbed. If this relation holds for the radiations of the X type as well as light, the results given above enable us to determine the frequency of the characteristic X radiations. In these cases w_0 is negligible compared with νh, so that

$$\nu = \frac{mv_s^2}{2h} = \frac{mk'^2 (Aw + B)^2}{2h} \quad \text{..............(32)}.$$

Thus for the copper-X rays belonging to the K series, putting

$$k' = 10^8,\ A = 1,\ w = 63,\ B = 0,\ m = \frac{1{\cdot}6}{1{\cdot}77} \times 10^{-27} \text{ and } h = 6{\cdot}55 \times 10^{-27},$$

$$\nu = 2{\cdot}74 \times 10^{18} \text{ sec.}^{-1}.$$

According to this estimate their frequency is about 2000 times as great as that of visible light.

X Rays and Crystals.

Since the foregoing account of the X rays was written, an enormous advance in our knowledge of their properties has taken place owing to a discovery made by Friedrich, Knipping and Laue*. These experimenters allowed a narrow beam of X rays to traverse a thin plate cut from a crystal of zinc-blende and then to fall on a photographic plate. When the plate was developed, instead of a single black spot being obtained as in the absence of the crystal, it was found that the central circular spot was surrounded by a regular pattern of circular and elliptical spots. This pattern varied with the direction in which the crystal plate was cut, and its symmetry was determined by the relation of the direction of the primary beam to the directions of the crystal axes. Laue, who appears to have suggested the experiment, explained the phenomena somewhat as follows:—The crystal is to be regarded as a geometrically regular arrangement of points (atoms) in space, each of which when traversed by the primary beam of X rays becomes a source of secondary waves. Owing to the regular arrangement of the points these secondary waves will only reinforce one another in certain directions. These directions determine the position of the secondary spots on the photographic plate. The theoretical investigation has been extended by Debye† so as to take account of the influence exerted by the heat motions of the atoms. Considering the case in which a system of atoms is arranged in cubical order so that the atoms lie on lines parallel to the axes of x, y and z at distance a apart, let a parallel beam of radiation of wave-length λ travel along the axis of x. The intensity of the radiation at a distant point x, y, z due to a small parallelepiped of such a substance at temperature T is proportional to

$$\frac{1}{r^2}\left(1+\frac{x^2}{r^2}\right)\left[N_1N_2N_3\left\{1-e^{-\frac{2RT}{f}\frac{4\pi^2}{\lambda^2}\left(1-\frac{x}{r}\right)}\right\}\right.$$
$$\left.+\frac{\sin^2 N_1\frac{\pi a}{\lambda}\left(1-\frac{x}{r}\right)\sin^2 N_2\frac{\pi a}{\lambda}\frac{y}{r}\sin^2 N_3\frac{\pi a}{\lambda}\frac{z}{r}}{\sin^2\frac{\pi a}{\lambda}\left(1-\frac{x}{r}\right)\sin^2\frac{\pi a}{\lambda}\frac{y}{r}\sin^2\frac{\pi a}{\lambda}\frac{z}{r}}\,e^{-\frac{2RT}{f}\frac{4\pi^2}{\lambda^2}\left(1-\frac{x}{r}\right)}\right] \quad \ldots\ldots\ldots(33),$$

---

* *Sitzber. der k. Bayer. Akad. d. Wiss.* June 1912.

† *Verh. d. Deutsch. Physik. Ges.* vol. xv. p. 678, p. 738 (1913).

where $r=(x^2+y^2+z^2)^{\frac{1}{2}}$, R is the molecular gas constant, f is the restoring force for unit displacement of the atoms, and $N_1 N_2 N_3$ are the number of atoms which lie along the x, y, z edges, respectively, of the parallelepiped. The bright spots are determined by the simultaneous vanishing of the three trigonometrical factors in the denominator

$$\sin^2 \frac{\pi a}{\lambda}\left(1-\frac{x}{r}\right) \sin^2 \frac{\pi a}{\lambda} \frac{y}{r} \sin^2 \frac{\pi a}{\lambda} \frac{z}{r}.$$

In addition to the diffraction pattern there is a uniform illumination proportional to

$$N_1 N_2 N_3 \{1-e^{-\frac{2RT}{f} \frac{4\pi^2}{\lambda^2}\left(1-\frac{x}{r}\right)}\}.$$

The relative intensities of the uniform illumination and the diffraction pattern depend upon the temperature T and on the stiffness of the atoms f. The results of the investigation are supported by the fact that in the case of diamond, for which f is very large, the diffraction pattern can be observed for much smaller values of x/r than for other substances. Formula (33) is deduced on the supposition that the heat motions of the atoms are in accordance with the Boltzmann-Maxwell law, and therefore requires modification in the light of recent work on the specific heats of solids. The necessary modifications are considered by Debye.

The production of diffraction effects with X rays is not confined to crystals. Thus Friedrich* has observed halos round the central image when X rays have been allowed to pass through certain non-crystalline substances.

It has been pointed out by W. L. Bragg† that the bright spots in the transmission photographs are found at places corresponding to regular reflexion from planes in the crystal rich in atoms, subject to the condition that the waves reflected from successive planes reinforce one another. This suggested that X rays would be reflected from the natural faces and cleavage planes of crystals, and the phenomenon was at once found‡ in the case of mica and other crystals. Since, for a particular wavelength λ of the X rays, there is no appreciable reflected intensity except in a particular direction determined by d/λ, where d is the distance between the successive layers of atoms, this phenomenon has opened up a new and powerful method of investigating the structure of crystals on the one hand and the properties of X rays and atomic structure on the other. These problems have been attacked with conspicuous success by W. H. and W. L. Bragg*, Moseley and Darwin†, Moseley‡ and others§. These investigations have confirmed in a remarkable way the conclusions as to the nature of X rays which we had already reached from less direct evidence.

---

* *Phys. Zeits.* Jahrg. 14, p. 1079 (1913).

† *Camb. Phil. Proc.* vol. XVII. p. 43 (1912).

‡ *Nature*, Jan. 23, 1913.

If the X radiation from an ordinary bulb, for instance one which is furnished with a platinum anticathode, is observed, it is found to consist of a general radiation similar to white light, except that the average wave-length is enormously shorter, accompanied by a number of sharp lines of much greater intensity than the neighbouring spectral distribution. In an X ray bulb with a platinum anticathode W. H. Bragg‖ found that the strongest line had a wave-length equal to $1{\cdot}10 \times 10^{-8}$ cm. and its mass absorption coefficient in aluminium was 23·7. According to Barkla's measurements this absorption coefficient corresponds to an X radiation of an element of atomic weight 74 if it is in the K series or 198 if it is in the L series. Since the atomic weight of platinum is 195 we are evidently dealing with the characteristic radiation belonging to the L series from this element. Admitting that the measured coefficient of absorption is slightly in error and that the true value in the L series would correspond to an atomic weight 195 the equivalent atomic weight in the K series is 72·5. According to Whiddington's results the energy of a cathode ray which will just excite the K radiation in an element of atomic weight 72·5 is $2{\cdot}06 \times 10^{-8}$ erg, whilst the energy $h\nu$ which corresponds to a wave-length $\lambda = 1{\cdot}10 \times 10^{-8}$ cm. is $1{\cdot}78 \times 10^{-8}$ erg. Thus the absolute numerical value of the frequency is in satisfactory agreement with the relations on p. 506. In addition to this platinum line Bragg also considers a line of wave-length $1{\cdot}66 \times 10^{-8}$ cm. emitted by a nickel anticathode and a line of wave-length $1{\cdot}25 \times 10^{-8}$ cm. emitted by a tungsten anticathode. The former belongs to the K series of nickel and the latter to the L series of tungsten. If these three lines were all in the K series the corresponding atomic weights of the elements which would emit them are 59, 67 and 72·5 respectively. The squares of these numbers are in the ratio 100 : 130 : 150 whilst the corresponding frequencies as determined by experiment are in the ratio 100 : 132 : 151. These results together form a striking confirmation of equation (32). The values of the wave-lengths given above have been deduced from the formula

---

* W. H. and W. L. Bragg, *Roy. Soc. Proc.* A, vol. LXXXVIII. p. 428 (1913); W. H. Bragg, *ibid.* vol. LXXXIX. p. 246 (1913); W. L. Bragg, *ibid.* p. 248 (1913); W. H. and W. L. Bragg, *ibid.* p. 277 (1913); W. H. Bragg, *ibid.* p. 430 (1914); W. L. Bragg, *ibid.* p. 468 (1914).

† *Phil. Mag.* June 1913; *ibid.* vol. XXVI. p. 210 (1913).

‡ *Phil. Mag.* vol. XXVI. p. 1024 (1913); *ibid.* vol. XXVII. p. 703 (1914).

§ De Broglie, *C. R.*, 17 Nov., 22 Dec. (1913), 19 Jan., 2 Feb., 2 Mar. (1914); Herveg, *Ver. d. Deutsch. Phys. Ges.* vol. XVI. p. 73 (1914).

‖ *Roy. Soc. Proc.* A, vol. LXXXIX. p. 246 (1913).

$$n\lambda = 2d\cos\theta \quad \text{......................(34)},$$

where n is the order of the spectrum, d is the distance between consecutive reflecting planes and θ is the angle of incidence. An exhaustive examination of the effects exhibited by rock-salt (NaCl) has shown that for reflecting planes parallel to the (1, 0, 0) face the value of d is $2{\cdot}81 \times 10^{-8}$ cm.

The X ray spectra of most of the elements whose atomic weights lie between those of aluminium (27·1) and gold (197·2) have been photographed by Moseley*. In the K series, which has been examined in the case of the elements between aluminium and silver (atomic weight 107·88) there are always two strong lines which are referred to as the α and β lines respectively. In the case of the elements of lower atomic weight, a number of faint lines are present in addition. In the L series, which has been examined in the case of elements ranging from zirconium (atomic weight 90·6) to gold there are usually five well-marked lines denoted by α, β, γ, δ and ϵ. These decrease in intensity as well as in wave-length from α to ϵ. In many cases a number of faint lines are also present. In both the K series and the L series there are simple numerical relations between all the lines belonging to any one of the sub-series α, β and so on from all the various elements.

Suppose we assign to each element an integral number N which, subject to two reservations, represents its position in a complete table of the elements arranged in sequence according to the atomic weights. In arranging this table the sequence of atomic weights is departed from when it clashes with the sequence of chemical properties as required by the periodic law. Thus argon is placed before potassium, cobalt before nickel and tellurium before iodine. In addition blank spaces for undiscovered elements are left between molybdenum and ruthenium, between neodymium and samarium, and between tungsten and osmium. N is called the atomic number of the element. Thus for hydrogen $N=1$, for aluminium $N=13$, for silver $N=47$ and so on. In accordance with the periodic law the chemical properties of the elements are determined by the atomic number N rather than by the atomic weight. Moseley finds that the frequencies ν of each sub-series are determined by the equation

---

* *Phil. Mag.* vol. XXVII. p. 703 (1914).

$$\nu = A\,(N-b)^2 \quad\text{......................}(35),$$

where A and b are constants characteristic of each sub-series. For the lines denoted by α in the K series

$$A = \left(\frac{1}{1^2}-\frac{1}{2^2}\right)\nu_0 \text{ and } b=1 \quad\text{..............}(36),$$

and for the lines denoted by α in the L series

$$A = \left(\frac{1}{2^2}-\frac{1}{3^2}\right)\nu_0 \text{ and } b=7{\cdot}4 \quad\text{...........}(37).$$

In (36) and (37) ν_0 is the fundamental Rydberg frequency in the formulae for the optical spectra of the elements. For the K series the agreement is quite exact, but for the L series there is a slight deviation from linearity with the elements of very high atomic weight. It is evident from these results that equation (32) is only an approximation, its validity depending on the fact that roughly speaking, the atomic weights of the elements are proportional to the values of N.

CHAPTER XX

SPECTROSCOPIC PHENOMENA

THE power of emitting light is a universal property of matter at a high temperature. When the matter is in a sufficiently concentrated condition the energy emitted within a given range of frequency $d\nu$ does not exhibit sharp variations from one wave-length or frequency to another. This is true of highly compressed gases as well as solids and liquids. The general character of the function which expresses the emitted energy of a given frequency in terms of temperature and frequency is similar to that of the corresponding function for black body radiation. In accordance with Stewart and Kirchhoff's law the difference between the emissivity and that of a black body depends only on the reflecting power; and when the latter is given as a function of ν and T the emissivity can be calculated from that of a black body.

The behaviour of gases which are not highly compressed is quite different. Almost the whole of the emitted energy is then confined to a limited number of quite narrow ranges of frequency. In consequence the gas is said to emit a line or a band spectrum according to the appearance of the light when examined in the spectroscope. The distinction between a line and a band spectrum is a sufficiently real one, but it is one which is difficult to define in simple terms. The width of the bright regions in the band spectra is greater than in the line spectra and the boundaries are less sharply marked. However, a number of so-called band spectra are found, under high resolving power, to consist of a multitude of fine lines very close together and it is possible that this may be a general feature of band spectra. Nevertheless, these two classes of spectra differentiate themselves quite sharply in other ways, as we shall see in a moment.

One of the most striking features of spectral lines is the constancy of their position in the spectrum. It is this fact which renders spectrum analysis so reliable. The only physical agencies which have been found capable of displacing them are intense magnetic fields (Zeeman effect), intense electric fields (Stark effect) and the application of high pressure. In all three cases the observed displacements are quite small. The intensities of different lines may be varied by any desired amount by changing the temperature and the manner and degree of excitation, without producing any change in the frequency. These frequencies are evidently characteristic of systems which remain identical under very varied physical conditions.

One is naturally tempted to try to account for spectral lines as the radiation from negative electrons vibrating about equilibrium configurations in the normal atom. Although there are many facts which seem to support such a view, the enormous complexity of the spectroscopic phenomena which have already been discovered is difficult to harmonize with such a simple hypothesis. In the spectra of a number of elements, iron for example, there are thousands of bright lines. In the case of iron, and many other elements, it has not been possible to find any simple relation between the spectral lines, but in a number of other cases many of the lines have been found to fall into series which exhibit fairly simple numerical relationships between the frequency numbers. What follows is the merest outline of the more important facts which have been discovered in this field. For further information the reader may be referred to J. Stark, *Principien der Atomdynamik*, vol. II. chap. II., and Kayser's *Handbuch der Spectroscopie*, vol. II. chap. VIII.

Series of Spectral Lines.

The first regularity of this kind was discovered by Balmer* in 1885 in the case of the line spectrum of hydrogen. The frequencies of the well-known lines H_α, H_β, H_γ, etc., are found to be given by the formula

$$\nu = \nu_0 \left(1 - \frac{4}{m^2}\right) \quad \text{......................(1),}$$

---

* *Verh. der Natur. For. Ges.* Basel, vol. VII. p. 548 (1885); *Ann. der Phys.* vol. XXV. p. 80 (1885).

where ν_0 is a constant and m takes in succession the integral values 3, 4, 5, etc. Starting from H_α, for which $\nu = \frac{5}{9}\nu_0$, the lines are widely spaced in the frequency scale at first, but the distance between each two succeeding lines continually diminishes. For large values of m all the lines practically coincide with $\nu = \nu_0$. The first 31 lines of this series have now been discovered and they all agree with the formula to within 0·6 Ångström unit (1 Ångström unit $= 10^{-8}$ cm. $=$ Å).

Balmer's series is much simpler than the series which have been found in the spectra of elements other than hydrogen. In the case of the alkali metals for example the series are made up of pairs of lines very close together (doublets). In the spectra of the alkaline earth and other metals the doublets are replaced by triplets. The individual lines in a doublet or triplet are called components. In a large number of cases 3 series and in some cases 4 series have been found in the spectrum of a single element. The separate series are also interrelated in an interesting manner. The values of the frequencies in the general case of 4 series of doublets are given by the following expressions.

Sharp Principal Series.

First Component, $${}_1\nu_1 = \frac{N_0}{(1+s)^2} - \frac{N_0}{(m+p_1)^2} \quad\ldots\ldots\ldots(2).$$

Second Component, $${}_1\nu_2 = \frac{N_0}{(1+s)^2} - \frac{N_0}{(m+p_2)^2} \quad\ldots\ldots\ldots(3).$$

Third Component, $${}_1\nu_3 = \frac{N_0}{(1+s)^2} - \frac{N_0}{(m+p_3)^2} \quad\ldots\ldots\ldots(4).$$

Sharp Subsidiary Series.

First Component, $${}_2\nu_1 = \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(m+s)^2} \quad\ldots\ldots\ldots(5).$$

Second Component, $${}_2\nu_2 = \frac{N_0}{(1+p_2)^2} - \frac{N_0}{(m+s)^2} \quad\ldots\ldots\ldots(6).$$

Third Component, $${}_2\nu_3 = \frac{N_0}{(1+p_3)^2} - \frac{N_0}{(m+s)^2} \quad\ldots\ldots\ldots(7).$$

Diffuse Principal Series.

$$\text{First Component,}\quad {}_3\nu_1 = \frac{N_0}{(1+d')^2} - \frac{N_0}{(m+p)^2} \quad \ldots\ldots(8).$$

$$\text{Second Component,}\quad {}_3\nu_2 = \frac{N_0}{(1+d'')^2} - \frac{N_0}{(m+p)^2} \quad \ldots\ldots(9).$$

$$\text{Third Component,}\quad {}_3\nu_3 = \frac{N_0}{(1+d''')^2} - \frac{N_0}{(m+p)^2} \quad \ldots(10).$$

Diffuse Subsidiary Series.

(In general this series consists of groups of six lines, divided unequally into 3 sub-groups, which we may term components.)

$$\begin{matrix}\text{First}\\ \text{Component}\end{matrix}\left\{\begin{aligned}
{}_4\nu_1' &= \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(m+d')^2} \text{ (strongest line)} \quad \ldots(11).\\
{}_4\nu_1'' &= \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(m+d'')^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(12).\\
{}_4\nu_1''' &= \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(m+d''')^2} \text{ (weakest line)} \ldots\ldots(13).
\end{aligned}\right.$$

$$\begin{matrix}\text{Second}\\ \text{Component}\end{matrix}\left\{\begin{aligned}
{}_4\nu_2'' &= \frac{N_0}{(1+p_2)^2} - \frac{N_0}{(m+d'')^2} \text{ (stronger line)} \ldots\ldots(14).\\
{}_4\nu_2''' &= \frac{N_0}{(1+p_2)^2} - \frac{N_0}{(m+d''')^2} \text{ (weaker line)} \ldots\ldots(15).
\end{aligned}\right.$$

$$\begin{matrix}\text{Third}\\ \text{Component}\end{matrix}\left\{ {}_4\nu_3''' = \frac{N_0}{(1+p_3)^2} - \frac{N_0}{(m+d''')^2} \ldots\ldots\ldots\ldots\ldots\ldots(16).\right.$$

In these formulae all the quantities on the right-hand side except m are constants. N_0 has the same value for all elements. The others vary from one element to another. The successive lines are obtained by giving to m the successive integral values 1, 2, 3, 4, etc. In the doublet series the frequencies denoted by ${}_1\nu_3$, ${}_2\nu_3$, ${}_3\nu_3$, ${}_4\nu_1'''$, ${}_4\nu_2'''$ and ${}_4\nu_3'''$ are missing and in the single line series the frequencies ${}_1\nu_2$, ${}_2\nu_2$, ${}_3\nu_2$, ${}_4\nu_1''$ and ${}_4\nu_2''$ in addition. For a critical discussion of the degree of accuracy with which formulae of this character fit the observed facts the reader may be referred to recent papers by W. M. Hicks in the *Philosophical Transactions of the Royal Society*.

The formulae given above show that, as in the hydrogen series, the lines all gradually crowd together as the frequency corresponding to $m = \infty$ is approached. This frequency, in the case of the first component for example, for each of the four series, has the values $\frac{N_0}{(1+s)^2}$, $\frac{N_0}{(1+p_1)^2}$, $\frac{N_0}{(1+d')^2}$ and $\frac{N_0}{(1+p_1)^2}$. Thus the two subsidiary series have the same termination. This is true separately of each corresponding pair of components. There are other relations between the different series.

In the case of the sharp subsidiary series and the diffuse principal series the difference in frequency between the components of a given line is constant throughout a given series. A similar relation holds between the lines whose frequencies are given by (13), (15) and (16) on the one hand and by (12) and (14) on the other in the diffuse subsidiary series. In the case of the first two components the values of the constant frequency differences in the respective series are, in the order of the table above,

$$\left.\begin{aligned} {}_2\nu_1 - {}_2\nu_2 &= \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(1+p_2)^2} \\ {}_3\nu_1 - {}_3\nu_2 &= \frac{N_0}{(1+d')^2} - \frac{N_0}{(1+d'')^2} \\ {}_4\nu_1'' - {}_4\nu_2'' = {}_4\nu_1''' - {}_4\nu_2''' &= \frac{N_0}{(1+p_1)^2} - \frac{N_0}{(1+p_2)^2} \end{aligned}\right\} \quad \ldots(17).$$

In the case of the sharp principal series on the other hand the difference in frequency between the components is not constant but decreases as the value of m increases. Thus the components all crowd together as the end of the series is approached. A similar relation holds between the lines given by equations (11)—(13) and (14) and (15) respectively in the diffuse subsidiary series. In these cases the series of individual components terminate at a common frequency; in the former cases there is a separate termination for each component of a given series.

The series are also interrelated in the following manner: The first term ($m = 1$) of the sharp principal series and the first term of the sharp subsidiary series have the same frequency; this is equal to the difference of frequency between the ends of the sharp principal series and of the sharp subsidiary series. Also the frequency difference between the ends of the diffuse principal series and the diffuse subsidiary series is equal, for each component, to the frequency of the first corresponding term of the diffuse subsidiary series.

Roughly speaking the intensity of the lines tends to diminish as m increases, although the relative intensities vary a great deal with the mode of excitation of the lines. In a sharp principal series of doublets the component of higher frequency is more intense than the slower component. The order of variation of intensity of the components with increasing frequency is usually the reverse of this in the case of the lines belonging to either of the subsidiary series.

The most important steps in the establishment of the serial relations just described are due to the work of Runge*, Rydberg†, Kayser, Paschen and Bergmann. These relations have been somewhat generalized recently by Ritz and Paschen and the results used, following a suggestion of Rydberg, to predict the existence of "combined" series. The success which has attended this development may be judged from papers by Paschen in the *Annalen der Physik* from 1908 to the present time‡.

The X-ray spectra of the elements have been described in the preceding chapter, p. 510.

Series in Band Spectra.

A number of relations between the emission frequencies observable in band spectra have been discovered by Deslandres§. The frequencies of the lines which make up a single band are found to agree very approximately with the formula

$$\nu = A\,(m + \alpha)^2 + c \quad \text{.....................} (18),$$

where A, α and c are constants and m takes various successive integral values. As a rule this formula does not deviate seriously from the observed values unless the band contains more than about 50 lines.

In general, emission bands are not found isolated but occur in related groups and in some cases there are a number of groups which exhibit relations with one another. Deslandres finds that the following relation covers the whole system of groups of bands in certain cases:

---

* *Brit. Assoc. Report for* 1888, p. 576.

† *K. Svenska. Vet. Akad. Handl.* vol. XXIII. No. 11, p. 155 (1890).

‡ See also Stark, *Principien der Atomdynamik*, vol. II. p. 50.

§ Papers in *Comptes Rendus* from 1885–1891 and from 1902–1905.

$$\nu = f(n^2, p^2) \times m^2 + Bn^2 + \phi(p^2) \quad \ldots\ldots\ldots\ldots(19).$$

Here ν is the frequency of a line in any of the bands, B is a constant, ϕ and f are characteristic functions, p, n and m are integers. The number p is characteristic of the group and n is characteristic of the band in which the line occurs, in the same way that m is characteristic of the particular line in the particular band. For specified values of p and n (19) becomes identical with (18) with $\alpha = 0$. The values of p and n never exceed 10, whereas m may be comparable with 100.

No explanation of (19) has ever been suggested; but it is important to observe that the numerical relation between the frequencies of lines in band spectra is of a character quite different from that which occurs in the series observable in line spectra. With regard to the foregoing system of series of line spectra it is also important to remember that very few of the lines in the more complex spectra have been found to fall into these series; so that the possibility of the existence of other types of series of lines, as well as of lines which do not belong to any series at all, is one which we have to bear in mind.

The Zeeman Effect.

We have several times had occasion to refer to the change in the frequency of spectral lines caused by the application of a magnetic field, which was discovered by Zeeman. In Chap. XVI we even considered the theory of this effect in connection with the theory of diamagnetism. We shall now discuss the theory in what is probably the simplest possible existing case.

Consider an electron which is subject to a restoring force proportional to its displacement from the equilibrium position in the atom. The restoring force per unit displacement is isotropic, that is to say it has the same value for different directions of displacement, in the absence of a magnetic field. Taking the equilibrium position as origin the equations of motion of the particle, when there is no magnetic field, may be written

$$m\frac{\partial^2 x}{\partial t^2} = -fx, \quad m\frac{\partial^2 y}{\partial t^2} = -fy, \quad m\frac{\partial^2 z}{\partial t^2} = -fz.$$

The natural frequency $n_0/2\pi$ has the same value whatever the direction of vibration and is given by

$$n_0^2 = f/m \quad \text{.........................(20)}.$$

Now suppose a magnetic field H_z parallel to the z axis is applied. As the electron is in motion this will add to the forces previously acting a new force proportional to the vector product of the velocity and H. The components of this new force on the electron are respectively

$$\frac{eH_z}{c}\frac{\partial y}{\partial t}, \quad -\frac{eH_z}{c}\frac{\partial x}{\partial t}, \quad 0 \quad \text{..................(21)}.$$

The equations of motion in the magnetic field are therefore

$$m\frac{\partial^2 x}{\partial t^2} = -fx + \frac{eH_z}{c}\frac{\partial y}{\partial t}, \quad m\frac{\partial^2 y}{\partial t^2} = -fy - \frac{eH_z}{c}\frac{\partial x}{\partial t}, \quad m\frac{\partial^2 z}{\partial t^2} = -fz \quad \text{.........(22)}.$$

These are solved by

$$x = a_1 \cos(n_1 t + b_1), \quad y = -a_1 \sin(n_1 t + b_1) \quad \text{......(23)},$$

$$x = a_2 \cos(n_2 t + b_2), \quad y = a_2 \sin(n_2 t + b_2) \quad \text{.........(24)},$$

$$z = a_3 \cos(n_0 t + b_3) \quad \text{.....................(25)},$$

where

$$n_1^2 - \frac{eH_z}{mc} n_1 = n_0^2 \quad \text{.....................(26)},$$

$$n_2^2 + \frac{eH_z}{mc} n_2 = n_0^2 \quad \text{.....................(27)}.$$

Equation (25) shows that the vibrations parallel to the magnetic field are unaffected by it. The corresponding frequency n_0 has the same value whether the field is applied or not. The six arbitrary constants a_1, a_2, a_3, b_1, b_2, b_3, of which the first three determine the amplitude and the last three the phase of the vibrations, depend on the initial conditions of motion and are of no particular interest in connection with the present discussion.

Equations (23) represent two rectilinear simple harmonic motions at right angles. These are of equal amplitude and differ in phase by a quarter period. They therefore constitute a circular vibration in a plane perpendicular to the magnetic field. The same is true of equations (24) except that the direction of the rotation is reversed. Since the difference between n_1 or n_2 and n_0

is very small in comparison with either, we can write, instead of (26) and (27),

$$\delta n_1 = n_1 - n_0 = \frac{eH_z}{2mc}, \quad \delta n_2 = n_2 - n_0 = -\frac{eH_z}{2mc} \quad \ldots(28).$$

Thus the difference between the frequency of each of the two circular vibrations and that of the original vibration has the same magnitude but is positive in the one case and negative in the other. The frequency difference per unit magnetic field is quite independent of the frequency of the original line and is entirely determined by the universal electronic constant e/m.

Now consider the radiation which will be emitted by an electron moving in the manner we are contemplating. Since the circular vibrations are capable of resolution into simple harmonic motions, the nature of the radiation, although not necessarily its amount, can be deduced from the simpler case of a rectilinear simple harmonic motion. By Poynting's theorem the rate of transmission of energy at any point is proportional to the vector product of the electric and magnetic intensities and the direction of transmission is the direction of that vector. It follows from the results of Chapter XII that in any motion in which the velocity and acceleration are collinear the magnetic intensity lies in circles about the axis of motion and vanishes at points along this axis. There is therefore no radiation along the direction of the axis of a simple harmonic motion. Now turn to the radiation in any of the directions which are perpendicular to this axis. It follows from Chapter XII that the electric intensity in the radiation wave in this case lies in the plane containing the axis of motion and the radius. It is also perpendicular to the latter. The magnetic intensity is equal to the electric intensity and its direction is normal both to the electric intensity and the radius. Since the plane of polarization of the radiation is that which contains the direction of the magnetic intensity in the wave-front, we see that the radiation emitted in the direction under consideration is completely polarized in the equatorial plane. These principles together with equations (23)—(25) are sufficient to determine completely the character of the radiation which is emitted in directions coincident with, and normal to, the lines of magnetic force.

Consider the light emitted along the lines of force first. The simple harmonic motion given by (25) emits nothing in this direction; so that the original line will be entirely absent when observations in this direction are made. It is different with the circular vibrations (23) and (24). The direction of observation is now situated in the instantaneous equatorial plane for these motions. The observed radiation will therefore have its electric intensity parallel to the instantaneous direction of motion and its magnetic intensity parallel to the radius drawn from the centre to the instantaneous position in the equivalent orbit. The light will therefore exhibit complete polarization. The direction of the rotation given by (23) is in the clockwise direction as seen from points along the positive z axis, that given by (24) is in the opposite direction. Thus the light emitted along the positive axis of H consists of two lines circularly polarized in opposite directions; the one which has the frequency n_1 is right-handed, the other which has the frequency n_2 is left-handed.

When the light is observed in a direction perpendicular to H_z the vibration corresponding to equation (25) will give rise to a line which is polarized in the plane to which the magnetic intensity is perpendicular. The frequency of this vibration is n_0, the same as that of the original line in the absence of a magnetic field. The circular vibrations are now observed in their own plane and may be resolved into simple harmonic motions along and perpendicular to the line of sight. The former give rise to no emission in the direction contemplated, whilst the latter give rise to light polarized in the plane containing the magnetic intensity. Thus when the emitted light is viewed in directions normal to the magnetic force the original line is converted into a triplet. The middle component has the same frequency as the original line and is polarized perpendicularly to the lines of magnetic force, whilst the other components are at equal distances from it and are polarized in the perpendicular plane.

According to (28) n_1 will exceed n_0 if e is positive and be less than n_0 if e is negative. Zeeman found that when the light was emitted along the positive z axis the slowest component of the observed doublet exhibited right-handed circular polarization. This shows that the centres of emission are *negatively* charged particles.

The problem of the effect of a magnetic field on electrons executing closed orbits in the atom is quite complex in detail;

but Larmor* has shown in a general manner that for small values of the magnetic force the motion consists in the combination of the original motion with a uniform rotation about the axis of H. This leads to equations (28) for the Zeeman shift.

All of the foregoing conclusions†, many of which were predicted by Lorentz, were found by Zeeman and other investigators to be accurately fulfilled in the case of a large number of spectral lines. According to Paschen all the lines of the single line series of spectra exhibit this relatively simple type of Zeeman effect. The same is true also of many other lines. Thus Purvis finds about 50 lines in the spectrum of palladium which give the normal effect. I ought also to add that the value of e/m deduced from the magnitude of the shift in these cases agrees with that given by the electrons furnished by the cathode rays and from other sources. It seems fairly clear from these results that the hypothesis of vibrating electrons is an important step towards the explanation of emission spectra.

In the case of the majority of spectral lines the Zeeman effect is more complicated than the "normal" type just outlined. But even the more complex cases exhibit certain relatively simple features which are of importance. For example, when the lines are observed in the direction of the magnetic field, the components of lower frequency exhibit right-handed, and those of higher frequency left-handed, polarization, showing that the vibrators are negatively charged. Moreover Runge‡ has shown that although the frequency displacement $\delta n/H$ often differs from the theoretical amount $e/2mc$, yet it is always a small integral multiple of $e/2mnc$, where n is a small integer. This holds true even when, as we shall see, the number of new lines produced by the magnetic field is much greater than two. Another important result, discovered by T. Preston§, is that all the lines of a given series and of homologous series of different elements are decomposed by the magnetic field in the same manner.

---

* *Aether and Matter*, p. 341.

† The rather different treatment considered in Chap. xvi, when it is fully worked out, leads to results identical with those above. Cf. Lorentz, *Theory of Electrons*, p. 124.

‡ *Phys. Zeits.* vol. viii. p. 232 (1907).

§ *Phil. Mag.* vol. xlv. p. 325 (1898), vol. xlvii. p. 165 (1899).

To illustrate this it will be best to consider one or two examples. The behaviour of the lines in the sharp principal series of doublets is exemplified by the case of the sodium D lines. Their behaviour in a transverse magnetic field is exhibited in Fig. 53. The letters p and n denote that the lines are polarized parallel and perpendicular to the lines of magnetic force, respectively.

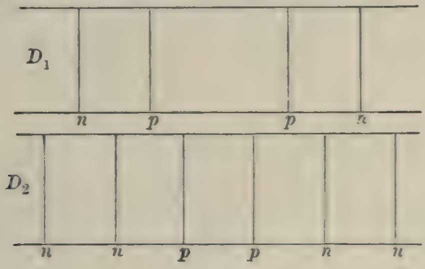


Fig. 53.

The two components of sharp subsidiary series of doublets are decomposed in exactly the same way except that the less refrangible component of the one series replaces the more refrangible of the other and vice versa. This may be regarded as a confirmation of the correspondence in the structure of the lines of these two series which is indicated by the respective series formulae.

The strong lines of the diffuse subsidiary series of doublets are split up into triplets, whilst the satellites split into eight components of which six are polarized perpendicularly and two parallel to the axis of the magnetic field.

The way in which the lines of the sharp subsidiary series of triplets break up in a transverse magnetic field is shown in Fig. 54. The behaviour of the diffuse series of triplets is still more complex, as is also that of many lines which have not been assigned to any series.

Paschen and Back* have recently observed that with certain double lines the character of the Zeeman effect depends very much on the strength of the magnetic field. The arrangement of the lines in strong fields is in some respects simpler than that in weak fields.

Not much has yet been accomplished in the way of an explanation of the more complicated types of Zeeman effect, although the theory has been attacked from several different points of view by Lorentz*, Voigt† and other writers. One important point brought out by Lorentz is that a magnetic field cannot of itself alone endow the vibrating atomic system with essentially new modes of vibration: it can only cause the separation of existing periods which previously were coalescent. This is an example of the well-known dynamical principle that the number of possible modes of motion of a system is determined by the number of degrees of freedom, since the number of degrees of freedom is not affected by the magnetic field. Naturally, by imagining the motion of the electrons to be constrained in various ways or by assuming different forms for the expressions for the potential and kinetic energies of the electrons, it is not difficult to arrive at rather complex types of Zeeman effect. But the theoretical results thus far attained do not seem to resemble very closely the effects exhibited by the actual spectral lines.

---

* *Ann. der Physik*, vol. xxxix. p. 897 (1912), vol. xl. p. 960 (1913).

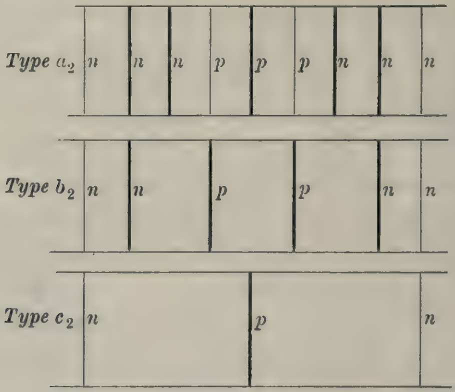


Fig. 54.

Most of the lines of band spectra which have been examined do not exhibit any measurable Zeeman shift. In the few cases in which this effect has been observed in the lines of band spectra the results have been more irregular than those given by line spectra. Thus Dufour‡ found in the banded flame spectra of the haloid salts of the alkaline earth metals some lines which, when observed in the direction of the magnetic field, gave rise to components with circular polarization in the opposite direction to that exhibited by the line spectra. Other lines showed the normal behaviour in this respect and in both cases the polarization was incomplete. The existence of circular polarization in the opposite direction to that given by the normal effect cannot necessarily be interpreted as implying positively charged vibrators, since Lorentz[*] has shown that this effect can arise from the vibrations of negative electrons under suitable circumstances.

---

* *Theory of Electrons*, chap. III.

† *Magneto- und Elektro-Optik*, chaps. II. and IV.

‡ *Comptes Rendus*, vol. XIV. pp. 118, 229 (1908).

Theories of Spectral Emission.

The frequency of the occurrence of the normal type of Zeeman effect and the fact that the more complicated types are closely related to the simple type and to the series of line spectra, show that the oscillations of negative electrons play a very important part in spectral emission. The chief difficulty which now lies in the way of the further development of the interpretation of these effects arises from our ignorance of the nature of the emitting systems. If we consider a particular series of lines, for example, we can either regard these as the overtones which accompany the fundamental vibration of a single system or we can look upon each line as the natural vibration of a separate system. In the latter case the other lines of the series are attributed to the occurrence of systems which are constitutionally related to one another in some regular way. Both of these views have their respective advocates. Another view, which attributed all the spectral lines to the natural modes of vibration of the normal atom, is certainly no longer tenable. As each electron can at most only give rise to three natural frequencies the number of electrons per atom which would be required to furnish the spectra of elements like iron and titanium is quite prohibitive. Moreover such a hypothesis would make the absorption spectra of metallic vapours quite different from what they are.

The experimenters who have adopted what one may call the overtone view of the nature of spectral series, have come to quite different conclusions as to the nature of the emitting systems. Thus Lenard[†] found that the upward stream of colour arising from a bead of alkali salt in the outer regions of a flame is undeflected by an electric field, whereas in the interior of the flame the coloration is deflected. In the former case only the lines of the principal series are emitted, whereas in the latter case lines of the subsidiary series may also be found. He therefore concluded that the emitters of the principal series were uncharged atoms of the metal and those of the subsidiary series atoms which had lost one or more electrons. Experiments with salt vapours in the arc confirmed this conclusion. On the other hand Stark*, largely from experiments made on the rapidly moving positive ions (canal rays) found in vacuum tubes under certain circumstances, has come to the conclusion that the emitters of all the line series are positive ions: and holds, on other grounds, that the band spectra are emitted by the neutral particles. There is no doubt that the sources of the line spectra are in many cases in rapid motion, since Stark† has shown that they exhibit the Doppler effect. This fact alone does not settle the question, but there is no doubt that the streams of deflected positively charged particles do in general give rise to the emission of series lines. However, such streams usually contain a fair proportion of neutralized particles which might be the source of the emission. In fact it seems to the writer that the bulk of the experimental evidence which has been brought to bear on this question might be interpreted in various ways.

---

[*] *Theory of Electrons*, p. 123.

[†] *Ann. der Physik*, vol. XI. p. 636 (1903), vol. XVII p. 195 (1905).

The truth of the overtone view of the nature of spectral lines cannot be said to have been established. In fact there are very grave objections to it. R. Ladenburg‡, who investigated the dispersion of luminous hydrogen in a vacuum tube, found that it was negligible except in the neighbourhood of the red line H_α. In this region satisfactory measurements showing the regular type of anomalous dispersion to be expected near a natural period were obtained, although no measurable effect could be detected in the neighbourhood of the bright blue-green line H_β. An extensive series of measurements covering the dispersion of the vapours of the different alkali metals in the neighbourhood of the lines of the various principal series has recently been carried out by Bevan§.

---

* *Jahr. der Radioakt. u. Elektronik*, vol. VIII. p. 231 (1911).

† *Phys. Zeits.* vol. VI. p. 892 (1905).

‡ *Ver. der Deutsch. Physik. Ges.* X. Jahrg. p. 858 (1908).

§ *Roy. Soc. Proc.* A, vol. LXXXIV. p. 209 (1910), vol. LXXXV. p. 54 (1911).

Both these investigations lead to important conclusions of a similar character.

Let us turn to the theory of dispersion given in Chap. VIII, p. 148. In the case of a gaseous substance we can neglect the term αP which depends upon the polarization, on account of its smallness, so that the refractive index, neglecting the effects specifically due to the absorption term in the equations of motion, will be given by

$$m^2 = 1 + \sum_1^n \frac{\nu_s e_s^2}{m_s (p_s^2 - p^2)} \quad \text{.................}(29),$$

where e is the charge, m the mass, of the vibrators, p_s the natural frequency of the sth type of vibrator, p the frequency of the light, ν_s the number of vibrators of the sth type present, and n the number of types. Bevan has shown that an equation like (29), with one term for each line, represents the measurements for potassium as accurately as they can be made; so that we shall not be led into any serious error by neglecting the absorption term. No doubt the same conclusions apply to the other substances investigated. Now in the case of the hydrogen lines e/m has the regular value for negative electrons, since the Zeeman effect is of the normal type. Thus all the quantities on the right-hand side of (29) are known except ν_s. It follows that from measurements of the refractive index we can obtain the number of emitting particles.

Now Ladenburg and Loria's experiments show that there is no appreciable dispersion in hydrogen until it becomes luminous, whereas luminous hydrogen shows anomalous dispersion in a marked manner in the neighbourhood of the line H_α. Thus it follows that the systems which emit this line *do not exist* in ordinary hydrogen but are only formed when the gas becomes luminous. Moreover luminous hydrogen does not show measurable anomalous dispersion near the line H_β, whence it follows that the number of systems which *can* emit light giving the line H_β is much smaller than the number which can emit light giving the line H_α. This conclusion must be valid even though we have neglected the absorption term; because it follows, from the shape of the dispersion curves, that the inclusion of the absorption term only affects the estimated value of ν_s seriously when we make use of frequencies very close to p_s. Now H_α and H_β are the first two lines of Balmer's series, so that this argument leads to the conclusion that the systems which can emit the different lines of this series are not present in equal numbers. It follows that the different lines of a given series are given out by different systems and therefore presumably by atoms in different states.

Bevan's results with the lines of the sharp principal series of the alkali metals confirm these conclusions. The argument here is not so strong because these lines do not exhibit the normal type of Zeeman effect, so that we have not such good grounds for the validity of the simple theory of dispersion which leads to (29). The constants might be of the more general type given in Chap. VIII, p. 176, for example. However, even in this case, it is not likely that the order of magnitude of ν_s estimated from the simple formula would prove to be seriously wrong.

As to the actual numerical differences in ν_s Bevan finds for the first four doublets of the principal series of potassium that the numbers of vibrating electrons are proportional to the respective numbers:

$$m_1 = 0{\cdot}113,\ m_2 = 1{\cdot}58 \times 10^{-4},\ m_3 = 1{\cdot}1 \times 10^{-5} \text{ and } m_4 = 3 \times 10^{-6}.$$

There is some evidence that the value of ν_s is equal, or approximately equal, for the two lines of a doublet. Results of a similar character are given by the other alkali metals. The proportion of systems capable of emission to the total number of atoms present is probably greatest in the case of the centres which emit the sodium D lines. Using data given by Wood for the temperature 644° C. Bevan estimates that the proportion of centres to atoms is about 1 in 12. A corresponding calculation has been carried out by Ladenburg and Loria in the case of hydrogen in a Geissler tube. By making use of the formulae (24 a) and (24 b) of Chap. VIII, in which the part played by the absorption term is taken into account, they conclude that about one centre capable of emitting the line H_a is found in every 50,000 molecules, under the conditions of their experiment.

As to the nature of the difference between the different atoms which makes them capable of emitting different spectral lines a number of plausible hypotheses may be considered. In the first place the electromagnetic analysis (measurement of the deflexion produced by transverse electric and magnetic fields) has shown that the atoms present in the vapours of metals may lose varying numbers of negative electrons. Thus Sir J. J. Thomson* has recently found that the positive rays in mercury vapour contain mercury atoms which have lost 1, 2, 3, 4, 5, 6 and 7 electrons respectively. In the second place there have also been shown to occur in vacuum tube discharges atoms which have gained one or more electrons in excess of the normal amount. Finally, systems may be formed by the combination of two or more atoms. I ought also to add that Ritz† has shown that if the electrons are supposed to vibrate in the field of a small magnet, different frequencies occur which obey the same kind of law as Rydberg's series, if the magnet is supposed to be built up of varying numbers of smaller elementary magnets. The possibility of this last type of theory has received some support from Weiss's work on the magnetic properties of bodies considered in Chap. XVI, but it does not otherwise seem to agree very well with present tendencies in the development of the theory of atomic structure (see Chap. XXI). H. A. Wilson‡ has pointed out that if we take the atom to be composed of a number of electrons in equilibrium inside a sphere of positive electrification of uniform volume density —a hypothesis which, as we shall see in the next chapter, gives a fair account of many of the properties of the atoms—then each atom possesses only one mode of vibration which is effective in producing any considerable amount of radiation, and the frequency of this mode of vibration depends only on the universal constants e, m and c, and the density ρ of the positive electrification in the sphere. According to this theory the frequencies of all spectral lines are determined by the density ρ of the positive electrification. For a given atom the requisite changes in ρ are secured either by a deficiency in the normal complement of electrons or by combination of the atoms with each other. As these changes always take place in discrete amounts the frequencies can be made to depend on whole numbers in the same way as in Balmer's series.

A number of attempts have been made to construct vibrating systems which have overtones resembling the spectral series.

---

* *Phil. Mag.* vol. XXIV. p. 668 (1912).

† *Gesammelte Werke.* Paris (1911).

‡ *Phil. Mag.* vol. XXIII. p. 660 (1912).

Ritz[*] has succeeded in constructing two-dimensional dynamical systems of this kind, whilst Whittaker[†] has expressed the form which the requisite energy functions must take in a very general manner.

A theory of spectral emission, due to Bohr, which depends on different principles from any of the foregoing will be considered in the next chapter, p. 585. This theory offers a quantitative explanation of Balmer's series and is able to overcome most of the difficulties we have so far encountered.

Fluorescence.

Many substances when illuminated, let us say by monochromatic light, are found to emit light of a different colour or frequency. This phenomenon is called fluorescence. Sir George Stokes, who made very important researches on the subject, concluded that the fluorescent light was invariably of lower frequency than the exciting light. This generalization has been found to be only approximately true. The first exceptions to it were noticed by Lommel[‡]. The more recent investigations of Nichols and Merritt[§] have shown that in the case of certain substances which they examined, the relative distribution of energy in the fluorescent spectrum does not depend very much on the frequency of the exciting light, and that a considerable proportion of the emitted energy may belong to higher frequencies than that of the exciting radiation.

Very often the fluorescent emission lasts for some time after the exciting radiation has been cut off. The phenomenon is then often termed phosphorescence. There does not seem to be any very sharp line of demarcation between fluorescence and phosphorescence, as the duration referred to may take almost any value from zero upwards with different substances and according to circumstances. However, in the case of liquid and gaseous bodies this duration has always been found to be too small to measure; so that they are strictly fluorescent according to the definition.

---

[*] *Loc. cit.*

[†] *Roy. Soc. Proc.* A, vol. LXXXV. p. 262 (1911).

[‡] *Wied. Ann.* vol. III. p. 113 (1878).

[§] *Phys. Rev.* vol. XIX. p. 18 (1904).

Fluorescence is very susceptible to changes in the physical condition of the bodies exhibiting it. Thus some substances are fluorescent only when dissolved in certain non-fluorescent liquids; whilst others, barium platino-cyanide for example, are fluorescent in the solid state and not when dissolved. The luminous paints are examples of an important class of substances whose behaviour has been investigated by Lenard and Klatt*. They find that these bodies, which are of a saline character, only exhibit phosphorescence when they contain three different chemical substances. Only small traces of two of the three need be present. These facts show that phosphorescence is not a unique property of the atom, or even of the chemical molecule.

It is probable that phosphorescence is invariably accompanied by the liberation of electrons†. This is a very important point and at once suggests the explanation of these phenomena. It now seems fairly certain that the exciting light first causes the emission of electrons from some constituent of the material and that the fluorescent light arises from the recombination of these emitted electrons. The time factor will then be determined by the resistance to this recombination and will be greater in solids than in liquids and gases, and at low than at high temperatures in solids. These requirements are borne out by the experimental results. In fact, all the observed phenomena seem to receive a very plausible explanation on this view.

It is questionable whether fluorescence can be sharply distinguished from other kinds of optical absorption except such as arise from electrical conduction. It is at least possible that in all cases energy is absorbed from the light by the electrons until disruption occurs. The disruption might occur entirely within an atom and not give rise to perceptible emission of electrons. The frequency of the emitted light might have any value including that of the incident light as a particular case. There might even be no re-emission of radiation except very indirectly, the whole of the absorbed energy being stored temporarily in the atom. In fact this hypothesis seems wide enough to cover all the phenomena referred to, and so far as the writer is able to judge, there are no facts known at present which definitely contradict it.

---

* *Ann. der Phys.* vol. xv. pp. 225, 425, 633 (1904).

† Stark and Steubing, *Phys. Zeits.* vol. IX. pp. 481, 661 (1908); Lenard and Saeland, *Ann. der Phys.* vol. XXVIII. p. 476 (1909).

Wiedemann and Schmidt* discovered that the vapours of various organic compounds and of the metals sodium and potassium exhibited fluorescence. This list has been extended to include the additional elementary substances mercury, iodine and bromine by R. W. Wood. The fluorescence exhibited by the vapours of the alkali metals and iodine has been investigated in detail by Wood† and found to exhibit remarkable results. Although there is a good deal of similarity in the effects exhibited by the different substances, perhaps the case of iodine is most interesting. When the stimulation is effected by monochromatic light the fluorescent spectrum is not a continuous one but is in the form of a series of fine lines whose frequencies are approximately equidistant, with the original line forming one of the lines of the series. The positions of the lines of the fluorescent spectrum change in a remarkable manner as the frequency of the exciting light is altered. They are also intimately related to the fine structure of the very complex absorption spectrum of iodine. This spectrum has seven very fine lines, visible only with the highest resolving power, within the width of the green mercury line. When the green mercury line is used for excitation it is found that a large number of the fluorescent lines are made up of fine lines having a structure similar to the seven absorption lines covered by the exciting spectrum, only the fluorescent lines are about thirty times as far apart on the frequency scale as the corresponding absorption lines. Many other interesting peculiarities have been observed which promise important results when investigated further.

The line fluorescence shown by the elements alluded to does not obey Stokes's law. It possesses the remarkable property‡ of being converted into a band spectrum by an admixture of inert gases like helium. The intensity of the banded fluorescence diminishes with the amount and molecular weight of the strange gas added. No adequate theory of these phenomena has yet been put forward, although it is clear that any satisfactory theory of spectroscopic phenomena will have to take account of them. The remarkable complexity of the absorption spectrum of iodine, which has been pointed out by Wood[*], should also be remembered in this connection.

---

* *Ann. der Phys.* vol. LVII. p. 447 (1896).

† Wood's *Physical Optics*, chap. XVIII., New York (1905); *Phil. Mag.* Oct. 1912.

‡ Wood and Franck, *Ver. der Deutsch. Phys. Ges.* vol. XIII. p. 84 (1911).

The Effect of Pressure on Spectral Lines.

In 1896 Humphreys and Mohler[†] observed that the spectral lines of various metals were displaced towards the red end of the spectrum when they originated in the arc at a high pressure. The shift was approximately proportional to the pressure. Humphreys[‡] explained this displacement as arising from the Zeeman effect of the intermolecular magnetic fields. The writer[§] pointed out that there should be a displacement, of the observed order of magnitude, arising from the forced vibrations in the atoms of the gas near the emitting centres. An explanation having a similar physical basis was given about the same time by Larmor[‖]. A theory which is not very different in principle from the last two has recently been put forward by G. H. Livens[¶], who, however, comes to the conclusion that the shift should be proportional to the concentration of the emitting particles and only to the pressure of the gas in so far as it influences the concentration. No doubt effects of the kind contemplated by Livens may occur under favourable circumstances, but it does not seem probable that they have much to do with the pressure shift under the conditions in which it has been observed. In the first place the experiments described above on the dispersion of light by luminous gases and by the vapours of the alkali metals indicate that in the case of many of the lines which exhibit the pressure shift it is probable that there are very few centres in a volume of the dimensions of the wave-length of light, so that Livens's analysis will not apply without modification. It then reduces to much the same thing as that given by the writer. In the second place the shift should vary enormously for spectral lines of the same series to accord with Bevan's results. Finally, such a view is hardly likely to lead even to so much consistency as has been recorded in measurements of the pressure shift, although, it must be admitted, this is far from being all that might be desired.

---

[*] *Phil. Mag.* Dec. 1912.

[†] *Astrophys. Journ.* vol. III. p. 144 (1896).

[‡] *Jahr. der Rad. u. Elektronik*, vol. v. p. 324 (1908).

[§] *Phil. Mag.* vol. XIV. p. 557 (1907).

[‖] *Astrophys. Journ.* vol. XXVI. p. 120 (1907).

[¶] *Phil. Mag.* vol. XXIV. p. 285 (1912).

The Stark Effect.

Stark* has recently observed interesting changes in a number of spectral lines when the emission takes place in a strong electric field. As in the case of the Zeeman effect the phenomena are different for different lines and the observed effects depend upon the geometrical relation between the direction of emission and the direction of the electric field. When the lines H_α and H_β are under observation in a direction perpendicular to the electric field, it is found that five lines appear instead of each original line. These five lines are symmetrical about the original line with which the central one is coincident. The two outer lines are much stronger than the other three and are polarized parallel to the electric field. The three inner lines are polarized perpendicularly to the electric field. When the radiation is observed in the direction of the electric field three evenly spaced lines appear. These have the same wave-length as the three lines which were polarized perpendicularly to the field in the former case and are now unpolarized.

Of the helium lines which have been examined those belonging to the sharp principal series and to the sharp secondary series exhibit no noticeable decomposition. The lines belonging to the diffuse subsidiary series break up into three or four lines of unequal intensity, asymmetrically spaced with respect to the original line. None of these lines are coincident with the original line and they are unpolarized. In the case of helium the lines observed in the direction of the electric field present the same characteristics as in the transverse direction.

For the three lines (H_β, H_γ and He $\lambda = 4026$ Å) for which the data are most accurate the displacement of the components is proportional to the strength of the electric field. The helium line of wave-length 4472 Å, which is split up very asymmetrically, appears to deviate from this law. With a field of 13,000 volt cm.$^{-1}$ the difference in the wave-lengths of the two outer components was $3{\cdot}6 \times 10^{-8}$ cm. for H_β and $5{\cdot}2 \times 10^{-8}$ cm. for H_γ. A number of lithium lines exhibit effects of the same order as those given by hydrogen and helium. Here again the sensitive lines are in the diffuse subsidiary series. The lines of the two sharp series exhibit only comparatively small displacements. The displacements shown by a large number of lines of calcium and mercury which have been examined are all small.

---

* *Sitzungsber. d. k. Preuss. Akad. d. Wiss.* Berlin, 1913, p. 932; *Ann. der Physik*, vol. XLIII. p. 965 (1914). Stark and Wendt, *ibid.* p. 983; Stark and Kirschbaum, *ibid.* p. 991, p. 1017.

An explanation of these effects has been put forward by Bohr[*]. He points out that the external electric field will change the amount of energy in the atoms in the different stationary states (see p. 585), and thus involve a change in the frequency of the radiation which is given out in changing from one stationary state to another. Other explanations have been offered by Garbasso[†], Gehrcke[†] and Schwarzschild[‡].

The Inverse Zeeman Effect.

In our treatment of absorption and dispersion in Chap. VIII we omitted to consider effects which may arise when an external magnetic field is present. This defect will now be remedied, but the treatment given will be very brief. Those who wish for fuller information about this and related questions may be referred to Lorentz's *Theory of Electrons*, chap. IV; Wood's *Physical Optics*, chap. XVII; and especially to Voigt's *Magneto- und Elektro-Optik*, *passim*.

If we wish to take account of the effect both of a magnetic field and of absorption, we cannot, in general, afford to neglect any of the terms in the equations of motion of the electrons numbered (1) in Chap. VIII. We shall, however, simplify matters by assuming that there is only one kind of electron per atom whose motion is of importance. Let us first consider the case when the external magnetic intensity R lies along the z axis, which is that of the direction of propagation of the light. The equations of motion, under these conditions, reduce to

---

[*] *Phil. Mag.* vol. XXVII. p. 512 (1914).

[†] *Phys. Zeits.* Feb. 1 (1914).

[‡] *Verh. d. Deutsch. Physik. Ges.* Jahrg. 16, p. 20 (1914).

$$m\frac{\partial^2 x}{\partial t^2} = e(E_x + \alpha P_x) - \kappa x - \beta\frac{\partial x}{\partial t} + \frac{e}{c}R\frac{\partial y}{\partial t} \quad \text{......(33)},$$

$$m\frac{\partial^2 y}{\partial t^2} = e(E_y + \alpha P_y) - \kappa y - \beta\frac{\partial y}{\partial t} - \frac{e}{c}R\frac{\partial x}{\partial t} \quad \text{......(34)},$$

$$m\frac{\partial^2 z}{\partial t^2} = e(E_z + \alpha P_z) - \kappa z - \beta\frac{\partial z}{\partial t} \quad \text{..................(35)},$$

where $\kappa = \lambda^{-1}$ of Chap. VIII. If there are N of the movable electrons in unit volume, the components of the polarization P are given by

$$P_x = Nex,\ P_y = Ney \text{ and } P_z = Nez \text{...........(36)}.$$

Let the waves propagated through the medium depend on the time through the factor $e^{\iota pt}$ only; then, by making use of (36), equations (33)—(35) may be replaced by

$$E_x = (\gamma + i\delta)P_x - i\epsilon P_y \text{ ....(37)},$$

$$E_y = (\gamma + i\delta)P_y + i\epsilon P_x \text{..................(38)},$$

$$E_z = (\gamma + i\delta)P_z \text{ ....(39)},$$

where $$\gamma = \frac{\kappa}{Ne^2} - \alpha - \frac{mp^2}{Ne^2},\quad \delta = \frac{\beta}{Ne^2}p,\quad \epsilon = \frac{R}{cNe}p \text{ ......(40)}.$$

Since the waves, supposed plane, are propagated along the direction of the z axis, the electric and magnetic vectors in the wave front will contain t and z through the factor $e^{\iota p(t-qz)}$ only, and will be independent of x and y. Thus the fundamental electromagnetic equations

$$\operatorname{rot} H = \frac{1}{c}\frac{\partial D}{\partial t}$$

and $$\operatorname{rot} E = -\frac{1}{c}\frac{\partial H}{\partial t}$$

give us

$$\left.\begin{aligned} c^2q^2E_x &= D_x = E_x + P_x \\ c^2q^2E_y &= D_y = E_y + P_y \end{aligned}\right\} \text{ ....(41)}.$$

Substituting these values for E_x, E_y in (37) and (38) we obtain

$$\left[\frac{1}{c^2q^2-1} - (\gamma + i\delta)\right]P_x = -i\epsilon P_y \text{...........(42)},$$

and $$\left[\frac{1}{c^2q^2-1} - (\gamma + i\delta)\right]P_y = +i\epsilon P_x \text{...........(43)}.$$

Thus $$P_y = \pm iP_x \text{ ....(44)}.$$

There are therefore two solutions having opposite values of P_x for a given value of P_y. To interpret the imaginary sign consider the case, corresponding to the solution having the positive sign, where P_x is proportional, at a fixed point, to the real part of $e^{\iota pt}$, or $\cos pt$. Then P_y will be proportional to the real part of $ie^{\iota pt}$, or $-\sin pt$. Thus the solution $P_y = +iP_x$ corresponds to a circular vibration of P, and since P is in the same direction as the resultant displacement of the electrons these also execute similar circular vibrations. The value of q for this vibration is given by substituting in either (42) or (43) from the equation $P_y = +iP_x$. It is thus given by

$$c^2q^2 = 1 + \frac{1}{\gamma + \epsilon + i\delta} \quad \text{..................(45).}$$

In a similar manner we see that the equation $P_y = -iP_x$ corresponds to a circular vibration in the opposite sense, for which q is given by

$$c^2q^2 = 1 + \frac{1}{\gamma - \epsilon + i\delta} \quad \text{..................(46).}$$

In general q will be complex. As in Chap. VIII, p. 165, we shall obtain the coefficient of absorption $\frac{k}{c}p$ and the index of refraction n if we put

$$cq = n(1 - ik) \text{........................(47),}$$

where n and k are real. n and k will thus be obtained by solving the simultaneous equations

$$n^2(1 - k^2) = 1 + \frac{\gamma \pm \epsilon}{(\gamma \pm \epsilon)^2 + \delta^2}, \qquad 2nk = \frac{\delta}{(\gamma \pm \epsilon)^2 + \delta^2} \text{...(48),}$$

where the positive sign corresponds to the vibration (45) and the negative sign to (46). When there is no magnetic field these equations become

$$n^2(1 - k^2) = 1 + \frac{\gamma}{\gamma^2 + \delta^2}, \qquad 2nk = \frac{\delta}{\gamma^2 + \delta^2} \quad \text{......(49).}$$

Equations (18) and (19) of Chap. VIII evidently reduce to (49) when the same assumptions are made and the slight difference in notation is allowed for.

This result establishes a simple correlation between the behaviour of the two circularly polarized rays in the longitudinal magnetic field and the equivalent unpolarized light in the absence of a magnetic field. The only difference between the equations (49) and the corresponding pair of equations (48) is that γ is replaced by $\gamma + \epsilon$ in the one case and by $\gamma - \epsilon$ in the other. Thus considering a given value of γ, the polarized ray (45), in the presence of the magnetic field, will have the same refractive index and absorption coefficient as the unpolarized ray for which γ had the value $\gamma + \epsilon$ in the absence of a magnetic field. Similar conclusions apply to the other ray if ϵ is replaced by $-\epsilon$. We see from (40) that p satisfies the relation

$$p^2 = \frac{Ne^2}{m}\left(\frac{\kappa}{Ne^2} - \alpha - \gamma\right),$$

so that the right-handed ray, which corresponds to (45), will exhibit the same behaviour as the unpolarized ray of frequency p shows in the absence of a magnetic field, provided its frequency p_1 is given by

$$p_1{}^2 = \frac{Ne^2}{m}\left(\frac{\kappa}{Ne^2} - \alpha - \gamma - \epsilon\right) = \frac{Ne^2}{m}\left(\frac{\kappa}{Ne^2} - \alpha - \gamma - \frac{R}{cNe}p_1\right) \ldots (50).$$

The left-handed ray (46) will exhibit the same behaviour as both the foregoing if its frequency p_2 satisfies

$$p_2{}^2 = \frac{Ne^2}{m}\left(\frac{\kappa}{Ne^2} - \alpha - \gamma + \frac{R}{cNe}p_2\right) \ldots\ldots\ldots\ldots (51).$$

Thus the curves that express the refractive index and absorption coefficient of the circularly polarized rays in the longitudinal magnetic field as functions of the frequency are separated by a frequency $p_2 - p_1$ given by

$$p_2{}^2 - p_1{}^2 = \frac{Ne^2}{m} \cdot \frac{R}{cNe}(p_2 + p_1),$$

or

$$p_2 - p_1 = \frac{e}{mc}R \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (52).$$

Each of these curves is identical with the corresponding curve in the absence of a magnetic field except for the displacement. The original curve is almost midway between the displaced curves. The relationship is exhibited graphically in Fig. 55, p. 544.

Comparing (52) with (28) we see that the displacement of the two circularly polarized components of the absorption spectrum is just equal to the Zeeman shift in the emission spectrum for the same magnetic field. Also if e is negative the absorption band for left-handed circularly polarized light is shifted towards the low frequency side. In the direct Zeeman effect we found that when e was negative the component of lower frequency was left-handed, if we observed the emission in the direction of the magnetic force.

Effects of the kind just specified were first observed by Maculoso and Corbino* in their experiments on the absorption of light by sodium vapour in the neighbourhood of the D lines, in the presence of a magnetic field. More complete experiments by Zeeman† and Hallo‡ have shown that the phenomena are completely in accordance with the theory, the development of which is largely due to Drude and Voigt. J. Becquerel§ has shown that similar phenomena are displayed by the very sharp absorption bands exhibited by the salts of certain rare earth metals at low temperatures, whilst Wood‖ has found them in the fine lines of the channelled absorption spectrum of sodium vapour. In both these cases the direction of rotation for some of the lines corresponds to that given by the elementary theory for positively charged particles (see p. 524).

We shall now turn to the case in which the light is propagated at right angles to the lines of force of the external magnetic field which we shall still suppose to lie along the z axis. The equations of motion of the electrons will therefore still be given by equations (33)—(35). Let the light be propagated along the x axis. The various electric and magnetic vectors will then depend upon the coordinates only through the factor $e^{\iota p(t-qx)}$. In the first place we notice that when the light is plane polarized so that the electric vibration is in the same direction as the magnetic field, the latter exerts no effect on the observed phenomena. For the relations between the polarization, the electric intensity and the motion of the electrons are given by equation (35) together with

$$D_z = P_z + E_z \quad \text{and} \quad P_z = Nez.$$

These equations, together with

$$\frac{\partial H_y}{\partial x} = \frac{1}{c}\frac{\partial D_z}{\partial t} \quad \text{and} \quad \frac{\partial E_z}{\partial x} = \frac{1}{c}\frac{\partial H_y}{\partial t},$$

---

* *Comptes Rendus*, vol. CXXVII. p. 548 (1898).

† *Amsterdam Proc.* vol. v. p. 41 (1902); *Arch. Néerl.* (2) vol. VII. p. 465 (1902).

‡ *Arch. Néerl.* (2) vol. x. p. 148 (1905).

§ *Comptes Rendus* (1906–7).

‖ *Phil. Mag.* vol. xv. p. 274 (1908).

are sufficient to determine completely the behaviour of the light, and none of them involve the external magnetic field R.

It is otherwise with the light polarized in the perpendicular direction. The effect of R has now to be considered, on account of equation (34). Since rot $H = \dot{D}/c$, we have

$$\operatorname{div} \dot{D} = c \operatorname{div} \operatorname{rot} H = 0,$$

and since the waves are plane

$$\frac{\partial D_y}{\partial y} = \frac{\partial D_z}{\partial z} = 0,$$

so that D_x will always have the value which it had before the light was present, namely zero. Thus

$$E_x + P_x = D_x = 0 \quad \text{.....................(53).}$$

From this relation, together with (37), (38) and $D_y = E_y + P_y$, we find

$$P_x = \frac{i\epsilon}{1+\gamma+i\delta} P_y \quad \text{..........................(54),}$$

$$P_y = \frac{1+\gamma+i\delta}{(\gamma+i\delta)(1+\gamma+i\delta)-\epsilon^2} E_y$$

$$D_y = \frac{(1+\gamma+i\delta)^2-\epsilon^2}{(\gamma+i\delta)(1+\gamma+i\delta)-\epsilon^2} E_y \quad \text{...........(55).}$$

Thus

$$c^2 q^2 = \frac{D_y}{E_y} = n^2(1-ik)^2 = \frac{(1+\gamma+i\delta)^2-\epsilon^2}{(\gamma+i\delta)(1+\gamma+i\delta)-\epsilon^2} \quad \text{...(56).}$$

From (56) the index of refraction n and the coefficient of absorption kp/c may be obtained by equating real and imaginary parts, and solving the resulting simultaneous equations for n and k.

From (54) we observe that the ratio of P_x to P_y is complex. If we turn this ratio into the form $ae^{2\pi i\theta}$, a will be the ratio of the amplitudes and θ the phase difference of the components of the polarization. Since P_x and P_y are proportional to x and y respectively, we see that in general the motion of the electrons will be in an ellipse perpendicular to the direction of the magnetic field. The eccentricity and orientation of the ellipse will evidently depend on the frequency of the light, since γ, δ and ϵ involve this quantity.

To determine the position of the absorption bands it is necessary to know the value of k. Solving the equations we find that this quantity is given by

$$k^2 = \frac{\sqrt{A^2 + B^2} - A}{C} \quad \text{.....................(57)},$$

where

$$A = [(1+\gamma)^2 - \delta^2 - \epsilon^2][\gamma(1+\gamma) - \delta^2 - \epsilon^2] + 2(1+\gamma)(1+2\gamma)\delta^2,$$

$$B = (1+2\gamma)\delta[(1+\gamma)^2 - \delta^2 - \epsilon^2] - 2(1+\gamma)\delta[\gamma(1+\gamma) - \delta^2 - \epsilon^2],$$

$$C = [\gamma(1+\gamma) + \delta^2 - \epsilon^2]^2 + (1+4\epsilon^2)\delta^2.$$

The exact determination of the maximum values of k from (57) by differentiation leads to formulae which are too complicated to handle, but it happens that these maxima are given very approximately by the minimum values of C. These values are readily found in the case of very sharp bands or absorption lines, where the region in which perceptible absorption occurs is so narrow that δ and ϵ may be treated as constants compared with γ, which passes through the value zero in this neighbourhood and varies very rapidly with p.

Subject to these approximations the minima of C and maxima of k are given by

$$\gamma(1+\gamma) = \epsilon^2 - \delta^2$$

or

$$\gamma = -\tfrac{1}{2} \pm \sqrt{\epsilon^2 - \delta^2 + \tfrac{1}{4}} \quad \text{..................(58)}.$$

The corresponding values of A, B, C and P_x/P_y are

$$A = 4\delta^2\epsilon^2, \quad B = \tfrac{1}{2}\delta\left(1 + 4\epsilon^2 \pm \sqrt{4\epsilon^2 - 4\delta^2 + 1}\right), \quad C = \delta^2(1+4\epsilon^2),$$

$$ae^{2\pi i\theta} = \frac{P_x}{P_y} = \frac{i\epsilon}{\tfrac{1}{2} \pm \sqrt{\epsilon^2 - \delta^2 + \tfrac{1}{4}} + i\delta} \quad \text{...........(59)}.$$

The simplest case arises when the absorption is relatively feeble and the magnetic field is strong enough to produce a separation of the absorption lines which is great compared with their width. Under these circumstances δ is large compared with unity (see Chap. VIII, p. 167) and ϵ must be large compared with δ; so that from (58) we have the approximate relation

$$\gamma = \pm\,\epsilon = \pm\,\frac{R}{cNe}\,p \quad \text{....................(60)}.$$

This shows that the original line is split into two absorption lines which are separated by the same frequency difference as the corresponding lines in the longitudinal inverse effect or the corresponding emission lines in the direct effect.

To the same order of accuracy the maximum coefficient of absorption is evidently, from (57), the same for both lines and is given by

$$\frac{kp}{c}=\frac{p}{4\delta c} \quad\text{..........................(61).}$$

This is just one-half of the value of the maximum absorption coefficient in the absence of a magnetic field under otherwise identical conditions (see Chap. VIII, p. 167). Thus when the medium is traversed perpendicularly to the direction of the magnetic force, by light in which the electric intensity is also perpendicular to the magnetic force, the light will be absorbed to the extent given by (61) when the frequency has either of the values given by (60). Corresponding to one of these vibrations the polarization of the medium and the motion of the electrons will be given respectively by $P_x=+iP_y$ and $x=+iy$, and corresponding to the other by $P_x=-iP_y$ and $x=-iy$. Thus the electrons move in opposite directions in circles perpendicular to the direction of the magnetic force. This motion agrees with that which corresponds to light of the same frequency when the inverse longitudinal Zeeman effect is under consideration. In general these transverse effects will not be symmetrical about the frequency corresponding to $\gamma=0$ because the expressions for n and k involve odd as well as even powers of γ.

Magnetic Rotation and Double Refraction.

The rotation of the plane of polarization of plane polarized light, when it passes through a refracting medium in the direction of the lines of force of an applied magnetic field, was discovered by Faraday* when he was engaged in trying to find the connection between electromagnetism and light. The explanation of this, the first magneto-optical effect to be discovered, follows at once from the theory of the longitudinal inverse Zeeman effect which has just been given.

As is well known, a plane polarized ray of light is equivalent to two rays circularly polarized in opposite senses and differing suitably in phase. Thus the plane polarized ray can be regarded as compounded of two opposite circularly polarized rays having the same frequency. But we have seen that these constituent rays are propagated with different speeds in the direction of the lines of magnetic force, so that when they emerge from the medium they will no longer have the same phase difference as when they entered. They will combine into a ray which is, in general, polarized in a plane different from that of the original ray. This rotation of the plane of polarization is easily calculated in the case in which the absorption coefficient kp/c is so small that we may neglect it. This condition must certainly hold when we are dealing with transparent refracting media.

---

* *Exp. Res.* § 2152 (1845).

Turning to equations (48) we see that when k is negligible, the refractive index is

$$n = 1 + \tfrac{1}{2}\frac{\gamma \pm \epsilon}{(\gamma \pm \epsilon)^2 + \delta^2} \quad \text{.....................(62)}$$

according as the ray is right- or left-handed. The velocities v_r and v_l of the right- and left-handed rays are therefore given by

$$\frac{1}{v_r} = \frac{1}{c} + \frac{1}{2c}\frac{\gamma + \epsilon}{(\gamma + \epsilon)^2 + \delta^2}$$

and

$$\frac{1}{v_l} = \frac{1}{c} + \frac{1}{2c}\frac{\gamma - \epsilon}{(\gamma - \epsilon)^2 + \delta^2}.$$

If one of the circularly polarized rays gets ahead of the other in phase by an angle ϕ the plane of polarization of the resultant ray will clearly be rotated through the angle $\phi/2$. The rotation ω of this plane per unit length of the medium, arising from the difference in velocity of the two rays, will thus be

$$\begin{aligned}\omega &= \frac{p}{2}\left[\frac{1}{v_r} - \frac{1}{v_l}\right]\\ &= \frac{p}{4c}\left[\frac{\gamma + \epsilon}{(\gamma + \epsilon)^2 + \delta^2} - \frac{\gamma - \epsilon}{(\gamma - \epsilon)^2 + \delta^2}\right] \quad \text{.........(63).}\end{aligned}$$

The sense of this rotation clearly reverses when the magnetic force R is reversed, because changing the sign of R replaces $+\epsilon$ by $-\epsilon$ and vice versa.

The rotation near an absorption line or band, where k cannot be disregarded, is most readily obtained graphically. Let $ABCDE$ be the curve which gives $1/v$ as a function of the frequency in the absence of a magnetic field. Then the corresponding quantity for the two circularly polarized constituents when a longitudinal magnetic field is present will be obtained, as is clear from the discussion on p. 538, if we simply displace $ABCDE$ to the right or to the left by an amount corresponding to $\pm \epsilon$. Let the curves thus displaced be $A_1B_1C_1D_1E_1$ and $A_2B_2C_2D_2E_2$ (Fig. 55). In these curves the ordinates represent $1/v$ and the abscissae the values of p which is 2π times the frequency. The value of the coefficient of rotation ω will be given as a function of the abscissae p by the lower curve $PQRST$ in which the ordinates are the differences of the ordinates of the two upper curves at the same value of p.

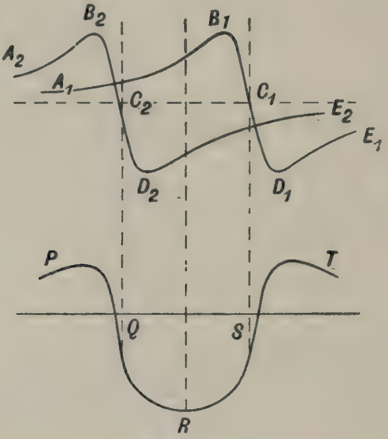


Fig. 55.

These conclusions hold good when the light is propagated along the lines of magnetic force. When the direction of propagation is transverse to the magnetic field a different phenomenon, which was predicted and verified by Voigt, manifests itself. Suppose the incident light to be plane polarized. We can then resolve it into two components one polarized in the plane containing the magnetic field and the axis of propagation, and the other in the perpendicular direction. In the latter component the electric vibrations are parallel to the magnetic intensity, so that the velocity of this constituent is the same as that of light of the same frequency in the absence of a magnetic field. This conclusion follows from the discussion on p. 539. The velocity of the perpendicular component is, however, given by the value of n

obtained from equation (56). The plane polarized ray therefore splits up into two plane polarized rays, polarized in mutually perpendicular planes, which travel with different velocities. Thus for light propagated transversely to the direction of the lines of magnetic force a simply refracting medium behaves, in a magnetic field, like a doubly refracting crystal. This effect does not reverse when R is reversed because (56) contains ϵ (and therefore R) only through its square.

Kerr's Magneto-optical Effect.

In 1876 Kerr[*] showed that when light, which is polarized in or perpendicular to the plane of incidence, is reflected from the poles of an electromagnet, the reflected light becomes elliptically polarized when the magnet is excited. To explain this effect on the electron theory it is necessary to take into account the modification of the laws of transmission and reflexion of light, which arises from the change in the motion of the electrons in the metal which is produced by a magnetic field. Naturally, the theory is more complicated than that of the inverse Zeeman effect. It is discussed at length by Voigt in *Magneto- und Elektro-Optik*, chaps. VI and VII.

Natural Rotatory Effects.

When a beam of plane polarized light passes through a plate of quartz so that the direction of propagation is parallel to the optic axis of the crystal, it is found that the emergent light is polarized in a different plane from that of the incident light. The rotation of the plane of polarization is proportional to the thickness of the plate and is somewhat different for light of different frequencies. The general character of the effect is thus similar to the Faraday effect, but a difference in the nature of the two effects is observable if the light is reflected so as to traverse the rotating system in the opposite direction. In the case of quartz the emergent light is then polarized in the original plane, whilst in the magnetic case the effect is equal to that produced by a path of double the length of the single path. Thus in the quartz plate the rotation relative to the axis of propagation is the same for both directions of propagation through the plate; whereas this is not so in the magnetic case, on account of the reversal of the direction of the magnetic field relative to the direction in which the light travels. It is found that some quartz crystals rotate the plane of polarization to the right and others to the left. The crystals themselves exhibit a similar geometrical dissymmetry, one being the mirror image of the other. The relation between them is like that between a left-handed and a right-handed helix.

---

[*] *B. A. Report*, 1876, p. 85; *Phil. Mag.* May, 1877.

Similar effects are shown by other crystals and by a very large number of liquids and solutions. In all the liquid substances which exhibit this effect, it has been found that the active constituent is either a compound of carbon in which at least one carbon atom is combined with four *different* chemical groups or a compound of some other element exhibiting an equivalent chemical structure. If this type of chemical structure is visualized in three dimensions it will be seen that there are only two possible ways of arranging the groups, so that the resulting molecules cannot be displaced in space so as to become coincident. These are so related that one of them is the mirror image of the other. There is thus the same relation between the molecules of these compounds as there is between the crystals of right- and left-handed quartz. Corresponding to the behaviour of quartz we should expect one of the two isomers, as they are called, to be left-handed and the other right-handed; otherwise, since the structural difference is purely a question of spacial arrangement, one would expect their other physical and chemical properties to be practically the same. Large numbers of such pairs of optical isomers have been isolated and their existence has led to great advances in our knowledge of structural chemistry since Pasteur*, Van 't Hoff†, and Le Bel‡ pointed out the importance of these phenomena.

It has been found that arrangements which rotate the plane of polarization of light may be constructed by taking a pile of thin sheets of doubly refracting crystals like mica, arranging them so that their optic axes are all similarly situated with respect to the axis of the pile and then shearing the system about the axis so that points which previously lay on a line parallel to the axis now lie on a spiral. It is possible that a sheared structure of this kind might account for the rotatory properties exhibited by some crystals, but it is obviously inadequate to account for the similar phenomena displayed by liquid substances. A general explanation based on the electron theory has been put forward by Drude*. It is only necessary to suppose that the electrons, which move with equal freedom in different directions on the usual form of dispersion theory (Chap. VIII), are constrained to move in helices in naturally active bodies. In that case Drude was able to show that right-handed and left-handed circularly polarized beams of light would travel with different velocities in the medium, the one being quicker than the other according to whether the helicoidal paths of the electrons are right- or left-handed. The rotation of the plane of polarization thus arises in a very similar manner to that considered in the inverse Zeeman effect and is greatest when the general dispersion is greatest. Taken as a whole the facts are in good general agreement with Drude's theory, but it is possible that other ways of introducing a helicoidal structure into the behaviour of the electrons would lead to very similar results.

---

* *Recherches sur la dissymétrie moléculaire des produits organiques naturels*, Paris (1860).

† *La Chimie dans l'Espace*, Rotterdam (1874).

‡ *Bull. Soc. Chim.* [2] vol. xxii. p. 377 (1874).

Electro-optics.

When various isotropic insulators are placed in a strong electric field they are found to behave optically like doubly refracting crystals. This effect, which is exhibited both by solids and liquids, was discovered by Kerr†. The difference in the refractive indices of the ordinary and extraordinary rays is proportional to the square of the applied electric intensity. In dealing with the theory of dispersion and related phenomena we made the assumption that when an electron is displaced from its equilibrium position in the atom it is acted on by a restoring force proportional to the displacement. Voigt‡ has developed a theory of Kerr's electro-optical effect which depends upon the supposition that this assumption is only true as a first approximation. When large forces and relatively large displacements are involved it is necessary to consider higher powers of the displacement than the first. In its general form this theory seems to be capable of embracing the known facts.

---

* *Lehrbuch der Optik*, chap. VI.

† *Phil. Mag.* [4] vol. L. pp. 337, 446 (1876); [5] vol. VIII. pp. 85, 229 (1879); vol. IX. p. 157 (1880); vol. XIII. pp. 153, 248 (1882).

‡ *Magneto- und Elektro-Optik*, chaps. VIII, IX and X.

Peculiar electro-optical effects are exhibited by many crystalline substances. Some of these appear to arise from the internal strains produced by an external electric field, an effect which is the converse of the piezo-electric effects investigated by J. and P. Curie*.

The effect of an electric field during the emission of spectral lines has already been considered (p. 534).

---

* *Comptes Rendus*, vol. xci. pp. 294, 383 (1880); vol. xcii. pp. 156, 350 (1881); vol. xciii. pp. 204, 1137 (1881).

CHAPTER XXI

THE STRUCTURE OF THE ATOM

THE way in which the electrons form comparatively stable groupings which exhibit the properties that characterize the atoms of the various chemical elements is a problem which has engaged the attention of a large number of physicists. It will be impossible to do more than briefly indicate some of the more interesting results which have been achieved in this field.

Perhaps the most striking property of the chemical atom is its definiteness and permanency. Its properties are only temporarily affected by the very strenuous actions which accompany chemical combination and decomposition. Sir Joseph Larmor* appears to have been the first to point out that this definiteness could not be expected to arise if matter consisted of nothing more than electrons of negligible linear dimensions whose motions were governed by the classical equations of the electrodynamic field. For consider any system in which the electric and magnetic vectors are given by the equations (1)—(4) of Chapter IX with $\rho = 0$. Let x, y, z and t be the space and time coordinates of any point of this system, E and H being the electric and magnetic vectors. Now consider a second system in which the corresponding variables are denoted by the suffix 1 and are such that

$$(x, y, z) = k(x_1, y_1, z_1), \qquad t = lt_1, \qquad E = mE_1, \qquad H = nH_1.$$

Then by substituting in the equations referred to, it is at once seen that, provided $k = l$ and $m = n$,

$$\operatorname{div} E_1 = 0, \qquad \operatorname{div} H_1 = 0,$$
$$\operatorname{rot} E_1 = -\dot{H}_1/c, \quad \operatorname{rot} H_1 = \dot{E}_1/c,$$

where the operations now refer to the subscripted independent variables. By introducing the condition that the charge of the electrons is to be the same in both cases we obtain an additional equation between k and m, but there is still nothing to determine the absolute value of these quantities. Thus, corresponding to any solution of the equations, there are an infinite number of other solutions corresponding to linear transformations of the variables. If any solution corresponds to a state of steady motion or rest there are an infinite number of such possible states. Thus the finite size and general definiteness of the atom must be due to something outside the equations referred to.

---

* *Aether and Matter*, p. 189.

This conclusion is valid only if the dimensions of the elements of electric charge are negligible compared with the dimensions of the problem. The radius of an electron, as deduced from its inertia, is about 10^{-13} cm. and is therefore about 10^{-5} times that of an atom. We should therefore expect that, as an approximation towards the problem of atomic structure, we may treat the negative electron as a point charge endowed with inertia. In that case there must be something which determines the finiteness and definiteness of the atom, which lies outside the properties of the negative electron itself as well as outside the equations of the field. It may therefore be expected to depend upon the properties of the positive electricity. One way in which this additional requirement has been met is by supposing the positive electrification to be limited to the region inside a sphere of atomic dimensions. This hypothesis was first introduced by Lord Kelvin[*]. Its consequences have been worked out very fully by Sir J. J. Thomson[†] and the question has also been considered by Lord Rayleigh[‡]. Other possibilities will be referred to later. (See pp. 580—587.)

A number of writers, in addition to those referred to above and in the preceding chapter, have considered the question of the emission of series of spectral lines by atoms made up of electrons. Among these are Jeans[§], Nagaoka[||] and Schott[¶]. One of the objects of these investigations has been to find systems of electrons which when disturbed from a state of equilibrium will give rise to a series of vibrations in which the overtones are related to the fundamental in the same way as the different members of Rydberg's series are. In this respect the researches referred to can hardly be considered to have been particularly successful, and it now seems as though this view of the relationship between the lines of spectral series must be a mistaken one (cf. Chap. XX, p. 525). On the other hand they have greatly added to our knowledge of the structural conditions which must be satisfied for systems of this kind to exist permanently.

---

[*] *Phil. Mag.* vol. III. p. 257 (1902).

[†] *Phil. Mag.* vol. VII. p. 237 (1904).

[‡] *Phil. Mag.* vol. XI. p. 117 (1906).

[§] *Phil. Mag.* vol. II. p. 421 (1901); vol. XI. p. 604 (1906).

[||] *Phil. Mag.* vol. VII. p. 445 (1904).

[¶] *Phil. Mag.* vol. XII. p. 21 (1906); vol. XIII. p. 189 (1907); vol. XV. p. 438 (1908).

The question of the definiteness of the atom, already alluded to, carries with it that of the definiteness of frequency, or the fineness, of spectral lines. This fineness is of a very high degree. Recent researches* have shown that in the case of many lines, practically the whole of the observable width is due to the Doppler effect arising from the thermal motion of the molecules. The actual emitters must therefore be instruments of very great precision. Moreover, as Jeans† has pointed out, they must be atoms, or at any rate systems whose symmetry is such that they are incapable of acquiring through molecular impacts any appreciable motion of rotation about any axis except that of the emitter. For instance if the system contains an emitting doublet whose axis is rotating with angular velocity $2\pi\omega$ about a fixed axis, the observed light will consist of three lines having frequencies $p+\omega$, p and $p-\omega$ instead of the single line of frequency p emitted by the system when at rest. If ω has the range of values given by the Boltzmann-Maxwell law of equipartition of energy, this effect would result in the production of a rather broad band fading away at the edges, instead of a line of the sharpness which is usually observed. In the case of monatomic molecules ω is inappreciable, as is shown by the specific heats of monatomic gases.

Disregarding the perturbation due to rotation it is necessary to consider how the frequency p of the atom at rest can be sufficiently definite. If the electrons are somehow arranged in positions of statical equilibrium, and the spectral lines are caused by vibrations about these positions, their definiteness is accounted for. On the other hand the existence of paramagnetism leads us to infer the presence, at least in certain atoms, of electrons rapidly revolving in closed orbits. Consequently numerous attempts have been made to seek the origin of spectral lines in the kinematics of revolving rings of electrons. Unless rather special and somewhat strange assumptions are made, however, it is impossible to secure sufficient definiteness of frequency in this way.

---

* Buisson and Fabry, *Journ. de Physique*, vol. II. p. 442 (1912).

† *Phil. Mag.* vol. II. p. 422 (1901).

The difficulty here, which has also been emphasized by Jeans, is brought out in the following discussion which is due to Schott*. Consider a system containing a ring whose radius is r consisting of n electrons, evenly spaced about the circumference when undisturbed, and revolving with velocity $V = \beta c$. Approximate stability is assured by the presence of a central positive charge or a containing sphere of positive volume electrification or some other device giving rise to an attraction towards the centre of the ring. The equations of motion are shown to be

$$nU = -\frac{r^2}{c^2}\frac{\partial (mV)}{\partial t} \quad \text{.....................(1)},$$

and
$$P_1 - \frac{c^2}{r^2}(1+\beta^2)K + \frac{e^2}{r^2}W^2 = \frac{mV^2}{r} \text{..............(2)},$$

where U is a function of n and β which is proportional to the rate of radiation from the ring†. The terms in K and W in (2) represent the force which arises from the electrons in the ring itself, other than the one whose motion is considered. K is a function of n only and W of n and β. P_1 is the central force arising from the rest of the atom and the term on the right is the centrifugal force. P_1 may be considered as a function of r into whose form it is not necessary to inquire. m the mass of an electron is of the form $\frac{e^2}{c^2 a}\psi(\beta)$, where a is its radius. Now nU, being proportional to the radiation, is always positive if the system radiates at all. Thus if a is constant equation (1) shows that β cannot be constant for a radiating system. When β is given r may be obtained from (2), so that the radius of the ring also will vary with the time, when the system radiates. Subject to the constancy of a, there can therefore be no state of strictly steady motion which is compatible with the emission of radiation. If a is allowed to vary with the time then (1) and (2) can obviously be satisfied by constant values of V or β and r and a state of definitely steady motion is possible. This, however, makes the energy radiated come from the internal energy of the electron and, so far as one can see at present, there is no other warrant for such a hypothesis. In any event any part of the radiation which does not arise from the expansion of the electron would, by its emission, cause changes in β and r. Thus the expanding electron might account for the sharpness of some lines, but it is difficult to see why practically all of them should be sharp on such a view.

---

* *Phil. Mag.* vol. XII. p. 21 (1906).

† Cf. the value of U in *Phil. Mag.* vol. XII. p. 22 (1906) and of R in vol. XIII. p. 194 (1907).

In constructing a theory of the atom it is not necessary either that spectral lines should originate with revolving electrons or that the lines of spectral series should correspond to harmonics. The existence of para- and ferro-magnetic substances does, however, appear to necessitate the existence of revolving electrons in some atoms. Such electrons may be so arranged that their radiation is practically zero. Equations (1) and (2) show that under these circumstances β and r may be constant. The deviation from steady motion is determined by the amount of radiation, so that an atom may have revolving rings of electrons in it and still be quite definite provided these do not radiate. This condition can be secured by placing a sufficient number of electrons in any one ring, as the radiation decreases with great rapidity as the number is increased. This result, which can easily be established by calculation, is also obvious physically, since the amount of radiation is determined by the lack of the corresponding static field. On the other hand, for the emission of spectral lines we require a system which may be unstable, as it is known not to be the normal atom, but which *must* radiate and *must* have a definite frequency. Power of radiating and definiteness of frequency are precisely the qualities which do not concur in systems of revolving rings of electrons.

If revolving systems occur in the ultimate parts of matter their axes must be determined by the structure of the matter and must turn in space when the matter as a whole is turned. Otherwise the properties of crystals and magnetized bodies would depend on their orientation in space*. When bodies whose atoms contain systems of electrons in orbital motion are turned, a gyrostatic couple ought to be experienced. The fact that such an effect has not been observed cannot, I think, be regarded as disproving the existence of such systems. The moment of momentum of the revolving electrons is so small that one could only expect to detect such an effect by carrying out special experiments of a delicate character, and this has not yet been done.

In the papers of Schott to which we have referred, the questions of the stability of, and the amount of radiation from, rings of revolving electrons are both considered in detail. In most of the calculations the only assumption made as to the nature of the force attracting the electrons to the centre of the atom is that it is of the usual electromagnetic type. Some of the results are therefore more general than the corresponding ones given by the investigation which we shall now describe.

The properties of the type of atom which is made up of coplanar rings of electrons inside a large sphere of uniform indeformable positive electrification have been worked out by Sir J. J. Thomson † and lead to very interesting results. If b is the radius of the large sphere and ρ the density of the electrification in it, then the force, acting on an electron whose charge is e at a distance a from the centre of the sphere and due to the positive charge of the latter, is $\frac{4}{3}\pi\rho ea$. If there are ν electrons in the neutral atom, the charge E in the positive sphere is

$$E = \nu e = 4\pi\rho b^3/3.$$

Hence, in terms of the number and charge of the electrons, the force they experience at a distance a from the centre due to the positive sphere is $\nu e^2 a/b^3$. In addition to this the electrons are acted upon by forces due to their mutual repulsions. If O is the centre of the sphere an electron at A will be acted upon by an electron at B with a force e^2/AB^2 and the radial component of this $= \frac{e^2}{AB^2}\cos OAB$. If $OA = OB$, i.e. if the electrons are in the same ring, this repulsive force is $e^2/4OA^2 \sin\frac{1}{2}AOB$. Hence, if we have n electrons arranged at equal angular distances $2\pi/n$ round the circumference of a circle, the radial repulsion on one of them due to the others is

---

* Cf. Jeans, *Phil. Mag.* vol. XI. p. 606 (1906).

† *Loc. cit.*, cf. also *The Corpuscular Theory of Matter*, chap. VI., New York, 1907.

$$\frac{e^2}{4a^2}\left(\operatorname{cosec}\pi/n+\operatorname{cosec}2\pi/n+\operatorname{cosec}3\pi/n+\ldots\ldots+\operatorname{cosec}\frac{(n-1)\pi}{n}\right).$$

Calling the sum within the bracket S_n we have, if the electrons are at rest under their mutual forces,

$$\frac{\nu e^2 a}{b^3}=\frac{e^2}{4a^2}S_n$$

or

$$\frac{a^3}{b^3}=\frac{S_n}{4\nu}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3).$$

This will determine a state of possible equilibrium if the electrons are at rest. If they are rotating round the ring we shall have to take account of the centrifugal force and the equilibrium condition becomes

$$\frac{\nu e^2 a}{b^3}=ma\omega^2+\frac{e^2}{4a^2}S_n\quad\ldots\ldots\ldots\ldots\ldots\ldots(4).$$

These equations only determine a state of possible equilibrium. They do not tell us whether the equilibrium is stable or unstable. For such a ring to be a possible part of a normal atom it is necessary that this arrangement should be stable, otherwise the ring would break up under the action of any external force. To find out whether the equilibrium is stable or not it is necessary to calculate the forces called into play by an infinitesimal displacement and to see whether its direction is such as to cause the displaced electron to move back to its original position. This must be done both for radial and for tangential displacements in the plane of the ring and for displacements perpendicular to the plane of the latter. Let us illustrate this by considering one or two simple cases.

In the case of a single electron it is evident that the centre of the sphere is the only position of equilibrium. The equilibrium in this case is evidently stable. In the case of two electrons it is evident by symmetry that the only equilibrium position is that in which they lie along the same diameter of the sphere. If they are at rest at a distance a from the centre the repulsive force is $\frac{e^2}{4a^2}$ and the attraction is $\frac{2e^2a}{b^3}$. Hence the total force acting on each electron, reckoned positive if directed towards the centre of the sphere, is

$$R = \frac{e^2}{4}\left(\frac{8a}{b^3} - \frac{1}{a^2}\right).$$

This vanishes if $a = \frac{1}{2}b$. If a becomes $a + da$, R increases by

$$dR = \frac{e^2}{4}\left(\frac{8}{b^3} + \frac{2}{a^3}\right) da.$$

As the quantity in brackets is positive when $b = 2a$, it follows that the equilibrium is stable as regards radial displacements.

The reaction called into play by a displacement in any direction perpendicular to the radius may be calculated as follows. Let the angular displacement from the equilibrium position be $d\theta$; then the linear displacement is $a d\theta$. The attraction of the positive sphere is still towards the centre, so that the tangential restoring force will be due entirely to the other electron. The force between the two electrons is repulsive and equal to $\frac{e^2}{4a^2}$ along the line joining them, if we neglect small quantities. The tangential component of this is $\frac{e^2}{4a^2} \sin \frac{1}{2} d\theta$ or $\frac{e^2}{8a^2} d\theta$ to the first order. This is directed towards the position of equilibrium of the displaced electron; so that the equilibrium is stable for lateral displacements. As this is true whatever the direction of the displacement, the arrangement of two electrons along a diameter, each at a point halfway between the centre and the circumference, is one which satisfies all the conditions for stability.

We shall now consider the general case. Take the centre of the atom as origin and let the position of an electron be given by the cylindrical coordinates r, θ and z. r is the projection of the radius on the plane of the undisturbed orbit, θ is the angle r makes with a fixed line in that plane and z is the displacement perpendicular to the plane of r and θ. The coordinates are supposed to undergo small oscillations about the steady values $r = a$, $\dot{\theta} = \omega$ and $z = 0$. If the suffix s refers to a particular electron we can therefore put $r_s = a + \rho_s$ and

$$\theta_s - \theta_{s-1} = \frac{2\pi}{n} + \phi_s - \phi_{s-1} \quad \dots\dots\dots\dots\dots\dots (5),$$

where n is the number of electrons in the ring, ρ_s and z_s are small compared with a, and ϕ_s and ϕ_{s-1} are small compared with $2\pi/n$.

The radial repulsion exerted by the sth electron on the pth is

$$R_{ps} = -e^2 \frac{\partial}{\partial r_p} [r_p^2 + r_s^2 - 2 r_p r_s \cos(\theta_s - \theta_p) + (z_p - z_s)^2]^{-\frac{1}{2}}$$
$$= \frac{e^2}{4a^2 \sin\psi} \left\{ 1 - \frac{\rho_p}{a} \left(\tfrac{3}{2} - \frac{1}{2\sin^2\psi} \right) - \frac{\rho_s}{a} \left(\tfrac{1}{2} + \frac{1}{2\sin^2\psi} \right) - \tfrac{1}{2} (\phi_s - \phi_p) \cot\psi \right\} \ldots(6),$$

where $\psi = (p - s)\dfrac{\pi}{n}$.

The tangential force Θ_{ps} tending to increase θ_p is

$$\Theta_{ps} = -\frac{e^2}{r_p} \frac{\partial}{\partial \theta_p} \left(\frac{1}{[r_p^2 + r_s^2 - 2 r_p r_s \cos(\theta_s - \theta_p) + (z_p - z_s)^2]^{\frac{1}{2}}} \right)$$
$$= \frac{e^2}{4a^2} \frac{\cos\psi}{\sin^2\psi} \left\{ 1 - \tfrac{3}{2} \frac{\rho_p}{a} - \tfrac{1}{2} \frac{\rho_s}{a} - (\phi_s - \phi_p)(\cot\psi + \tfrac{1}{2}\tan\psi) \right\} \quad (7).$$

The force perpendicular to the plane of the undisturbed orbit is

$$Z_{ps} = \frac{e^2}{8a^3 \sin^3\psi} (z_p - z_s) \quad \ldots\ldots\ldots\ldots\ldots\ldots(8).$$

In equations (6)—(8) higher powers than the first of the small quantities ρ, ϕ and z have been dropped.

The radial force R_p exerted by all the other electrons in the ring on the pth is

$$R_p = \frac{e^2}{4a^2} S - \rho_p A' - \Sigma_s \rho_{p+s} A_s - a \Sigma_s \phi_{p+s} B_s \quad \ldots\ldots(9).$$

The tangential and perpendicular forces on the same particle are respectively

$$\Theta_p = \Sigma_s \rho_{p+s} B_s - a \phi_p C + a \Sigma_s \phi_{p+s} C_s \quad \ldots\ldots\ldots(10),$$

and

$$Z_p = z_p D - \Sigma_s z_{p+s} D_s \quad \ldots\ldots\ldots\ldots\ldots\ldots(11),$$

where

$$\left.\begin{aligned}
&S = \sum_{s=1}^{n-1} \left(1 \Big/ \sin\frac{s\pi}{n} \right), \quad A' = \frac{e^2}{4a^3} \left(\tfrac{3}{2} S - \tfrac{1}{2} \sum_{s=1}^{n-1} \left[1 \Big/ \sin^3 \frac{s\pi}{n} \right] \right), \\
&A_s = \frac{e^2}{8a^3} \left(\frac{1}{\sin\frac{s\pi}{n}} + \frac{1}{\sin^3\frac{s\pi}{n}} \right), \quad B_s = \frac{e^2}{8a^3} \left(\cos\frac{s\pi}{n} \Big/ \sin^2\frac{s\pi}{n} \right), \\
&C_s = \frac{e^2}{4a^3} \frac{\cos\frac{s\pi}{n}}{\sin^2\frac{s\pi}{n}} \left(\cot\frac{s\pi}{n} + \tfrac{1}{2}\tan\frac{s\pi}{n} \right), \quad C = \sum_{s=1}^{n-1} C_s, \\
&D_s = \frac{e^2}{8a^3} \frac{1}{\sin^3\frac{s\pi}{n}}, \text{ and } D = \sum_{s=1}^{n-1} D_s.
\end{aligned}\right\} (12).$$

The equations of motion of the pth electron are

$$m\left\{\frac{\partial^2 r_p}{\partial t^2} - r_p\left(\frac{\partial \theta_p}{\partial t}\right)^2\right\} = -\frac{ve^2 r_p}{b^3} + R_p \quad \text{.........(13)},$$

$$m\left(r\frac{\partial^2 \theta_p}{\partial t^2} + 2\frac{\partial r_p}{\partial t}\frac{\partial \theta_p}{\partial t}\right) = \Theta_p \quad \text{.....................(14)},$$

$$m\frac{\partial^2 z_p}{\partial t^2} = -\frac{ve^2}{b^3} z_p + Z_p \quad \text{.........(15)}.$$

From (13) since ω is the value of $\frac{\partial \theta}{\partial t}$ when the steady motion is undisturbed

$$\frac{ve^2 a}{b^3} = ma\omega^2 + \frac{e^2}{4a^2}S,$$

in agreement with (4). By making use of this relation, (13) becomes, to the first order in the small displacements,

$$m\frac{\partial^2 \rho_p}{\partial t^2} - 2ma\omega\frac{\partial \phi_p}{\partial t} = \rho_p\left(m\omega^2 - \frac{ve^2}{b^3}\right) + R_p - \frac{e^2}{4a^2}S \text{...(16)}.$$

By similar treatment (14) may also be made linear, to the same degree of approximation. Thus the values of ρ and ϕ are the solutions of a set of $2n$ simultaneous linear differential equations of the second order. To find the frequencies assume that ρ_p and ϕ_p vary as e^{iqt}. Equation (16) then becomes

$$(A - mq^2)\rho_p + A_1\rho_{p+1} + A_2\rho_{p+2} + \dots$$
$$- 2ma\omega iq\phi_p + aB_1\phi_{p+1} + aB_2\phi_{p+2} + \dots = 0 \text{...(17)},$$

where $A = \frac{e^2}{4a^3}S + A'$. Treated in a similar way (14) becomes

$$2m\omega iq\frac{\rho_p}{a} - B_1\frac{\rho_{p+1}}{a} - B_2\frac{\rho_{p+2}}{a} + \dots$$
$$+ (C - mq^2)\phi_p - C_1\phi_{p+1} - C_2\phi_{p+2} - \dots = 0 \text{...(18)}.$$

There are n each of the equations (17) and (18) obtained by giving p the successive integral values 1, 2, 3, ... n.

These equations can be solved by the following artifice. If $\rho_{p+1} = \alpha\rho_p$ and $\phi_{p+1} = \alpha\phi_p$ for all values of p from 1 to $n-1$, and if $\rho_1 = \alpha\rho_n$ and $\phi_1 = \alpha\phi_n$, then $\alpha^n = 1$ identically; so that α is one of the nth roots of unity and is equal to

$$\cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n},$$

where k is an integer between 0 and $n-1$. Whatever value p may have, we see that equations (17) and (18) reduce to

$$\rho_p(A - mq^2 + \alpha A_1 + \alpha^2 A_2 + \dots + \alpha^{n-1} A_{n-1}) \\ + \phi_p a(-2im\omega q + \alpha B_1 + \alpha^2 B_2 + \dots + \alpha^{n-1} B_{n-1}) = 0 \dots (19),$$

and

$$\rho_p(2im\omega q - \alpha B_1 - \alpha^2 B_2 - \dots - \alpha^{n-1} B_{n-1}) \\ + \phi_p a(C - mq^2 - \alpha C_1 - \alpha^2 C_2 - \dots - \alpha^{n-1} C_{n-1}) = 0 \dots (20).$$

Thus we see that all the $2n$ equations (17) and (18) are satisfied by $\rho_p = \alpha^p \rho_n$ and $\phi_p = \alpha^p \phi_n$ for all the values of p between 1 and n, provided (19) and (20) are also satisfied. By eliminating ϕ_p and ρ_p from (19) and (20) we see that the frequencies q are given by the biquadratic equation

$$(A - mq^2 + \alpha A_1 + \alpha^2 A_2 + \dots + \alpha^{n-1} A_{n-1}) \\ (C - mq^2 - \alpha C_1 - \alpha^2 C_2 - \dots - \alpha^{n-1} C_{n-1}) \\ = -(-2im\omega q + \alpha B_1 + \alpha^2 B_2 + \dots + \alpha^{n-1} B_{n-1})^2 \dots (21).$$

This may be written more concisely

$$\left(\tfrac{3}{4}\frac{e^2}{a^3} S + L_k - L_0 - mq^2\right)(N_0 - N_k - mq^2) = (M_k - 2m\omega q)^2 \dots (22),$$

where

$$\left.\begin{aligned} L_k &= \frac{e^2}{8a^3}\sum_{s=1}^{n-1} \cos\frac{2sk\pi}{n}\left[\frac{1}{\sin\frac{s\pi}{n}} + \frac{1}{\sin^3\frac{s\pi}{n}}\right] \\ N_k &= \frac{e^2}{4a^3}\sum_{s=1}^{n-1} \cos\frac{2sk\pi}{n}\,\frac{\cos\frac{s\pi}{n}}{\sin^2\frac{s\pi}{n}}\left(\cot\frac{s\pi}{n} + \tfrac{1}{2}\tan\frac{s\pi}{n}\right) \\ M_k &= \frac{e^2}{8a^3}\sum_{s=1}^{n-1} \sin\frac{2sk\pi}{n}\,\frac{\cos\frac{s\pi}{n}}{\sin^2\frac{s\pi}{n}} \end{aligned}\right\} \dots (23).$$

In these equations k is the integer in $\alpha = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}$ and may have any value from 0 to $n-1$. If $n-k$ is written for k in (22) the values of q given by the equations differ only in sign, so that the frequencies are the same. Thus all the values of the frequency can be got by putting $k = 0, 1, 2, \dots (n-1)/2$ if n is odd, and $k = 0, 1, 2, \dots n/2$ if n is even. If n is odd there are $(n-1)/2$ equations of the type (22) leading to independent values of the frequency. When $k=0$, $M_k=0$ and equation (22) reduces to a quadratic, so that altogether there are $4(n+1)/2-2=2n$ values of the frequency. When n is even there are $n/2+1$ equations, but $M_k=0$ when $k=0$ or $n/2$, so that two of these reduce to quadratics. Thus we see that whether n be odd or even there are $2n$ possible frequencies for vibrations in the plane of the orbit, corresponding to the $2n$ degrees of freedom of the electrons, in this plane.

In a similar way the n frequencies q of the vibrations at right angles to the plane of the orbit may be shown to be given by

$$\frac{\nu e^2}{b^3}+P_k-P_0-mq^2=0 \quad\text{.................}(24),$$

where $$P_k=\frac{e^2}{8a^3}\sum_{s=1}^{n-1}\cos\frac{2sk\pi}{n}\Big/\sin^3\frac{s\pi}{n}\quad\text{..............}(25).$$

The values of L, M, N and P have been worked out in a number of the simpler cases and the corresponding frequencies calculated*.

The importance of this investigation lies, for reasons which have already been given, not so much in the fact that it enables us to evaluate the frequencies of the vibrations of the electrons, as in the fact that the frequency equations (22) and (24) enable us to determine whether the equilibrium is stable or not. If all the values of q given by (22) and (24) are real then all the disturbances ρ, ϕ and z are periodic functions of the time, and the system will only execute small oscillations about the steady configuration. In this case the steady motion is stable. If, on the other hand, any of the roots contain an imaginary part, the corresponding values of ρ, ϕ or z will contain factors of the type e^{ht}, where h is a real positive constant. Thus any such disturbance will increase indefinitely with the lapse of time and the presence of complex or imaginary values of q shows that the equilibrium is unstable. The condition that the equilibrium should be stable is therefore that all the values of q given by the equations (22) and (24) should be real for all the admissible values of k.

---

* Cf. Thomson, *loc. cit.*

The stability of the z motions can always be provided for, since the value of the rotation ω is, so far, at our disposal. Transposing equation (9) we have

$$\frac{e^2}{4a^3} = \frac{1}{S}\left\{\frac{\nu e^2}{b^3} - m\omega^2\right\},$$

and, substituting this value of a, (24) becomes

$$mq^2 = \frac{\nu e^2}{b^3}\left\{1 + \frac{T_k - T_0}{2S}\right\} - m\omega^2\,\frac{T_k - T_0}{2S} \;\;\ldots\ldots\ldots(26),$$

where $$T_k = \frac{8a^3}{e^2}\,P_k \text{ and } T_0 = \frac{8a^3}{e^2}\,P_0.$$

$T_k - T_0$ is always negative, so that the right-hand side of (24) can obviously be made positive by choosing ω large enough. Thus q can always be made real and the z motions stable.

Now turn to the motions in the plane of the orbit. We may write equation (22) in the form

$$(A - q^2)(B - q^2) - (C - Dq)^2 = f(q) = 0 \;\;\ldots\ldots\ldots(27),$$

where

$$mA = \tfrac{3}{4}\frac{e^2}{a^3}S + L_k - L_0,\;\; mB = N_0 - N_k,\;\; mC = M_k \text{ and } D = 2\omega.$$

A, B, C and D are real and B is positive. We can make C/D as small as we please by suitably choosing ω, and at the same time preserve the stability of the z motions. Let A have any positive value, then C/D can always be chosen so as to be less than $+A^{\frac{1}{2}}$ or $+B^{\frac{1}{2}}$. Now $f(q)$ is positive when $q = \pm\infty$, negative when q lies between $+A^{\frac{1}{2}}$ and $+B^{\frac{1}{2}}$ or $-A^{\frac{1}{2}}$ and $-B^{\frac{1}{2}}$, and positive when $q = C/D$. The graph of $f(q)$ thus crosses the real x axis four times between $\pm\infty$. If A is negative there are only two such intersections. For all the roots to be real it is therefore necessary and sufficient that A should be positive. This is the condition for stability.

When the number of electrons in a ring becomes considerable, negative values of A begin to appear; so that no amount of rotation can make the ring stable. The greatest number of electrons which are stable in a circle inside a sphere, containing an equal total quantity of positive electricity uniformly distributed, is five[*]. Six or more electrons in a single ring are never in stable equilibrium however great the angular velocity is. Such a ring may, however, be made stable if a negative charge is introduced at the centre. If this negative charge is equal to that carried by p electrons, the effect of the additional repulsion thus introduced is to increase A in the frequency equation (27) to $A+3\frac{pe^2}{a^3}$[*]. This can always be made positive by taking p sufficiently large.

From what has been said it is clear that the condition for a ring to be stable is that $A+3\frac{pe^2}{a^3}=\frac{3}{4}\frac{e^2S}{a^2}-(L_0-L_k)+3\frac{pe^2}{a^3}$ should be positive. The greatest value of L_0-L_k is for $k=n/2$ when n is even and for $k=(n-1)/2$ when n is odd. The number of electrons which would have to be placed at the centre of the ring to ensure stability is therefore given by the least integral value of p for which

$$\left.\begin{array}{ll} & 3\frac{pe^2}{a^3}>L_0-L_{\frac{n}{2}}-\frac{3}{4}\frac{e^2S}{a^3}\text{ when } n \text{ is even} \\ \text{and} & 2\frac{pe^2}{a^3}>L_0-L_{\frac{n-1}{2}}-\frac{3}{4}\frac{e^2S}{a^3}\text{ when } n \text{ is odd}\end{array}\right\} \quad \ldots (28).$$

By means of these equations the least value of p for a given value of n can be calculated. p increases very rapidly for large values of n as is shown by the following table of corresponding values:

n	5	6	7	8	9	10	15	20	30	40
p	0	1	1	1	2	3	15	39	101	232

When p exceeds unity the additional electrons cannot all lie at the centre of the ring but must separate under their mutual repulsion. Thus when $n=9$ and $p=2$ there will be an external ring of seven and two electrons inside situated along a diameter of the sphere at equal distances from the centre. When p is large it is possible for the electrons to arrange themselves in stable equilibrium in rings rotating about a common axis and lying in parallel planes. On the assumption that the effect of external rings of electrons can be neglected and that the effect of internal rings is the same as if the total charge on the electrons were collected at the centre of the sphere, the number of electrons in the successive rings can be calculated by the application of the foregoing principles. It is probable that this treatment will give a fair representation of the properties of atoms containing a number of electrons, even if the latter are not arranged in rings but, for example, in concentric shells. The conditions for equilibrium are of the same general character in both cases. If a shell contains a large number of electrons it will be unstable unless some of them are inside, just as the corresponding ring was. We shall now consider the structure and properties of the atoms which have different numbers of electrons, when their constitution is determined by the equilibrium conditions to which we have been led. The numerical computations will be omitted; they are given in the paper by Thomson to which reference has frequently been made.

---

[*] Cf. Thomson, *loc. cit.*

We have seen that when there is only one electron it will lie at the centre of the positive sphere, whilst if there are two, they lie always along a diameter at equal distances from the centre. Three electrons will arrange themselves at the corners of an equilateral triangle situated on a circle whose radius is given by equation (3). The arrangement of four electrons at the corners of a square is unstable if the electrons are at rest. It is found that one of the electrons will jump out of the plane of the others and that the stable arrangement for four electrons is at the corners of a regular tetrahedron. A similar occurrence takes place whenever the number of electrons is greater than three. Thus three is the greatest number of electrons that can exist in stable equilibrium in a ring, provided they are at rest. By assigning to the rings of electrons sufficiently high velocities, it appears that the number in the ring can be increased to as many as five without instability setting in. A ring of six or any greater number, however, is unstable even if rotating. Six electrons can be in stable equilibrium if one is at the centre of the sphere and the other five rotate in a ring around it. A single electron at the centre is sufficient to make a ring of seven or eight stable, thus accounting for nine electrons altogether. With ten electrons it is necessary to have two in the centre and a ring of eight outside. With eleven three go to the centre and this holds till we get to fourteen, when the number in the centre increases to four. With seventeen electrons it is found that the stable arrangement is an outer ring of eleven with six inside it. We have seen, however, that a ring of six is unstable, so that one of them goes to the centre leaving a ring of five. We have thus three concentric systems in this case, containing respectively one, five and eleven electrons. When the number of electrons in the outer ring becomes considerable, the number which it is necessary to place inside the ring in order to ensure stability increases very rapidly. For large values of the number n in the ring the number p inside varies as n^3.

The actual arrangement of the electrons with some of the smaller numbers is given in the following table. The number of horizontal columns gives the number of rings; the numbers are the number of electrons in each ring. The total number in the atom is thus obtained by adding up the vertical columns.

outer ring	1	2	3	4	5

outer ring	5	6	7	8	8	8	9	10	10	10	11
inner ring	1	1	1	1	2	3	3	3	4	5	5

outer ring	11	11	11	12	12	12	12	13	13	13	13	14	14	15	15
2nd ring	5	6	7	7	8	8	8	8	9	10	10	10	10	10	11
1st ring	1	1	1	1	1	2	3	3	3	3	4	4	5	5	5

outer ring	15	15	15	16	16	16	16	16	16	16	17	17	17	17	17	17	17
3rd ring	11	11	11	11	12	12	12	13	13	13	13	13	13	14	14	15	15
2nd ring	5	6	7	7	7	8	8	8	8	9	9	10	10	10	10	10	11
1st ring	1	1	1	1	1	1	2	2	3	3	3	3	4	4	5	5	5

outer ring	17	……	19	20	20	20	20	20	20	20	20	21	21
4th ring	15	……	16	16	16	16	17	17	17	17	17	17	17
3rd ring	11	……	13	13	13	13	13	13	14	14	15	15	15
2nd ring	5	……	8	8	8	9	9	10	10	10	10	10	11
1st ring	1	……	2	2	3	3	3	3	4	4	5	5	5

outer ring	21	…………	22	…………	23	…………	24
5th ring	17	…………	20	…………	20	…………	21
4th ring	15	…………	16	…………	17	…………	17
3rd ring	11	…………	13	…………	13	…………	15
2nd ring	5	…………	8	…………	10	…………	11
1st ring	1	…………	2	…………	3	…………	5

Thus 93 electrons arrange themselves in six concentric rings of 5, 11, 15, 17, 21 and 24 respectively. 94 electrons begin the series with seven rings having the arrangement 1, 5, 11, 15, 17, 21 and 24.

Numbers like those in the preceding table showing the arrangement of the electron in the atom can be obtained in the following manner. It is first necessary to determine a sufficient number of corresponding values of p and n. This can be done by means of equations (28). Having done this we can draw the curve which gives p as a function of n. It is of the general form exhibited by the accompanying figure.

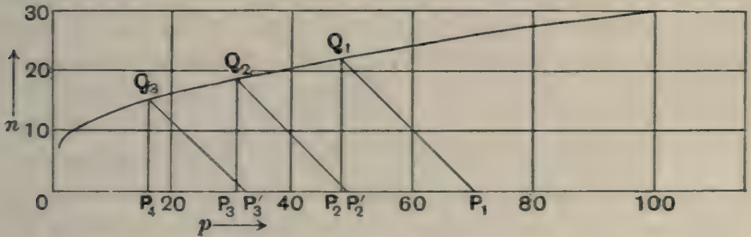


Fig. 56.

Suppose it is required to find how a large number N of electrons will arrange themselves in stable rings. Take $OP_1 = N$ along the axis of p and draw P_1Q_1 inclined at 135° to OP_1 intersecting the curve in Q_1. Then if Q_1P_2 is perpendicular to OP_1, $Q_1P_2 = P_2P_1$. By the property of the curve Q_1P_2 electrons in a ring require OP_2 inside to make them stable, so that $N = OP_1$ electrons will arrange themselves with an outer ring of P_2P_1 surrounding OP_2 electrons. In general OP_2 will not be an integer; in this case, instead of OP_2 we take OP_2' the next higher integral value. Thus we see that $N = OP_1$ electrons will arrange themselves so that the outermost ring contains $P_2'P_1$ electrons. We can find the distribution of the OP_2' electrons by repeating the process. From P_2' draw $P_2'Q_2$ inclined at 135° to OP_1 and draw Q_2P_3 perpendicular to OP_1. If OP_3' is the next integer higher than OP_3 the OP_2' electrons will arrange themselves with an outer ring of $P_3'P_2'$ electrons surrounding the remaining OP_3'. In this way we can proceed until the whole N electrons are accounted for.

Referring to the table on the preceding page we see that the successive atoms formed in this way possess features analogous to those properties of the chemical elements which are summarized by the periodic law. Thus it will be observed that the first element with seven rings is the same as that with six except for the outer ring of 24 electrons; this in turn is the same as that with five except for the outer ring of 21, and so on. We might expect that elements with the same internal structure would have similar properties. If so, we should expect them to be separated by groups of elements having widely different properties. If, as a number of different lines of investigation lead us to believe, the atomic weight is proportional to the number of contained electrons, this is exactly what happens. As we proceed through a list of elements of increasing atomic weight we find elements of similar character periodically appearing separated by entirely different elements. This kind of relationship ought to be exhibited especially by properties such as the frequency of vibration of the electrical constituents of the atom, which we should expect to be determined largely by the geometrical arrangement of the contained electrons. It is well known that the lines of the spectral series of elements of the same chemical family are closely related.

Another kind of resemblance to the properties of the chemical elements is brought out in a still more marked way if we can consider the equilibrium of the electrons in the successive artificial atoms and confine ourselves to the case where there is a constant number, for example twenty, of electrons in the outer rings. Starting with the first member, that with 59 electrons altogether, this will only just have enough electrons inside to keep the outer ring stable. It will therefore very readily give off one electron. When it has given this off, however, there will only be 58 electrons left, a number which is not great enough to have an outer ring of twenty. These will therefore arrange themselves with an outer ring of nineteen. Now 58 is the greatest number which can have an outer ring of nineteen, so that the stability of this atom as regards emission of electrons will be very high, more particularly as it has an excess of positive charge. The atom with 59 electrons will thus be capable of emitting one negative electron and so form a monovalent positive ion. It will do this with great readiness but will only be able to emit one negative ion. It will therefore behave like the strongly electropositive elements hydrogen, lithium, sodium, potassium, etc. The next atom with 60 electrons will be somewhat more stable than the first, but will be able to emit two electrons before the number falls to 58 and it becomes exceedingly stable as regards further emission. Thus it will resemble the divalent electropositive elements of the alkaline earth group. In a similar way the next element will not so easily emit an electron but it will be able to part with three before reaching the very stable condition with 58 electrons. It will therefore be less electropositive but will be capable of giving rise to a trivalent positive ion. It will therefore resemble the trivalent earth metals such as aluminium. In this way we see that as the number of electrons increases, the elements which have an outer ring of twenty become continuously less electropositive but have a continuously increasing electropositive valency. When we come to the last atom with twenty electrons in the outer ring and 67 altogether this would theoretically be able to emit nine electrons, but practically it will be so stable that it will be incapable of emitting any except under very great forces. It may thus be considered to resemble the inert gases helium, argon, etc. which are incapable of entering into chemical combination. Again, this element will be unable to combine with a free electron; for if it did so it would have 68 electrons altogether and these would arrange themselves with an outer ring of 21. As the system with 68 electrons and an outer ring of 21 is very unstable and liable to emit an electron when neutral, it will be still more unstable when it carries an excess of negative charge, so that this atom with 67 electrons will be incapable of combining with one additional electron. It will thus have zero electronegative valency. In this respect it again resembles the inert gaseous elements. The element with 66 electrons will tend to combine with one electron; since, when the atoms are neutral, the element with 67 is more stable than that with 66. It will not be able to combine with more than one, for if it did it would possess 68 electrons altogether, an arrangement which, as we have seen, is exceedingly unstable even when neutral. This element will thus behave like the strongly electronegative monovalent elements fluorine, chlorine, etc. The atom with 65 electrons will have a less strongly marked tendency to combine with an additional one but will be able to combine with two altogether before reaching the limiting condition. It will thus resemble the elements oxygen, sulphur, etc. which are divalent but less strongly electronegative than the preceding elements. Proceeding in this way we see that the element with the greatest number of electrons, out of those which have an outer ring of twenty, has zero electronegative valency. As the number of electrons is successively diminished the valency steadily increases whilst the elements become successively less electronegative. The most electronegative element has one less than the maximum number of electrons and is monovalent.

These properties furnish a striking analogy to the variation of valency and electrochemical properties as we pass through a series of elements in the periodic table. Starting with the element possessing 58 electrons, the maximum number with an outer ring of nineteen, the following table shows how the electropositive and electronegative valencies change as the number is increased:

Number in outer ring	19	20	20	20	20	20	20	20	20	20	21
Total number	58	59	60	61	62	63	64	65	66	67	68
Electropositive valency		1	2	3	4	5	6	7	8	9	1
Electronegative „	0	8	7	6	5	4	3	2	1	0	

The corresponding valencies for the series of elements between neon and argon, omitting potassium, are shown in the following table:

Element	Ne	Na	Mg	Al	Si	P	S	Cl	Ar
Atomic weight ...	20·2	23·0	24·3	27·1	28·3	31·0	32·1	35·5	39·9
Electropositive valency	0	1	2	3	4	5	6	7	0
Electronegative valency	0	7	6	5	4	3	2	1	0

In each case the sum of the maximum electropositive and electronegative valencies is constant, and the elements become successively more electronegative and less electropositive as the mass of the atoms increases.

The numbers purporting to represent the electropositive and the electronegative valencies of the different chemical elements are not in every case verified by facts. They are rather to be taken as representing a law which summarizes a general tendency. Thus so far as the writer is aware no compound of sodium is known in which it is heptavalent. This is probably due to the fact that it is difficult to get strongly electropositive elements to combine with one another. The electronegative elements are more adaptable in this respect, and they furnish numerous examples in which the law is fulfilled. We have for instance the following compounds which exhibit the maximum electropositive and electronegative valencies:

Electropositive valency	$SiCl_4$	PCl_5	SCl_6	Cl_2O_7
Electronegative „	SiH_4	PH_3	SH_2	HCl
		PNa_3	Na_2S	$NaCl$

In each case the sum of the valencies is constant and equal to eight.

The numerical relations obtained above are to be taken as illustrative rather than as indicating the actual number of electrons which the atoms of the respective chemical elements contain. If the electrons were arranged in shells instead of in rings, the conditions of stability would lead to results of the same general character, but the numerical values might be quite different. In order that the periodicity in the electrochemical properties and the observed changes in the electropositive and the electronegative valency should be found, all that is necessary is that there should be a sudden increase in the stability of the electrons in atoms which have certain numbers N_1, N_2, N_3, etc. of electrons and a gradual diminution as we pass from any one of these numbers to the next higher. For example, an atom which contains N_2+n electrons would lose n electrons rather easily, but it would require forces of a different order of magnitude to dislodge $n+1$ electrons. Its electropositive valency would thus be n. On the other hand it would have a tendency to combine with any number of electrons until the total number became equal to N_3. Its electronegative valency would therefore be $N_3-(N_2+n)$. The sum of the two valencies has the same value for all the elements with numbers of electrons between N_2 and N_3 and is equal to N_3-N_2. In discussing the question of valency it is not necessary to consider specifically the effect of the excess of positive or negative charge when electrons are removed from, or added to, a given atom. This effect will change regularly with the excess or deficiency of electrons. It is also necessary to observe that the question of valency is one of equilibrium in a system, the molecule, which is uncharged as a whole. The effect of the total charge on the individual atoms being different will therefore only come in to a very limited extent, since they are so very near together. It would be rather different if we were discussing the possibility of the ionization of a given atom.

It is worth while remarking that the elements which lie between the very strongly electronegative monovalent elements like fluorine and the next higher strongly electropositive monovalent element, e.g. sodium, are the gases He, Ne, Ar, Kr and Xe which have no chemical affinity; whereas those which lie between elements like manganese, which show some analogy with the chlorine group, and the corresponding element like copper, which is analogous to sodium, exhibit the highest electropositive valency of any of the elements. This is shown by the oxy-compounds and various complex amino-compounds of the metals of the iron, palladium and platinum groups.

It is convenient to have a name for the measure of the tendency of an electron to leave an atom which has been discussed above. Following Sir J. J. Thomson we shall refer to this as the electronic pressure of the atom.

Chemical Combination.

It has been pointed out that Lord Kelvin, in his paper entitled "Aepinus atomized" (*Phil. Mag.* vol. III. p. 257, 1902), was the first to suggest that the chemical atom consists of a sphere of uniform positive electrification containing negative electrons of much smaller dimensions embedded in it. In that paper he discusses the forces which will come into play when some of the simpler types of atoms are brought together. Considering the simplest type of all, that which contains only one electron, it is evident that two such atoms will exert no mutual force if the spheres lie entirely outside one another. If, however, one of the spheres A penetrates another B, then, since part of the positive sphere of B lies inside that of A, the repulsion of the positive sphere of B by that of A will be less than the attraction of the negative electron at the centre of A. Thus the two spheres will attract one another. It is clear, however, that the negative electron at the centre of each sphere will still be in equilibrium there until the centre of one sphere lies within the circumference of the other. If both the spheres are equal, the mutual repulsions of the negative electrons then exceed their attractions by the positive spheres; the electrons therefore move along the line joining the centres of the two atoms so as to lie outside the centres, but remain always within the atom. We thus get a neutral (uncharged) combination of the two atoms which may be regarded as the simplest type of elementary molecule. If the sizes of the spheres are unequal, the story is somewhat different. A calculation of the forces shows* that, whereas the electron in the smaller sphere moves outwards along the line of centres, that in the larger sphere moves towards the smaller. This movement towards the smaller sphere is gradual at first, but at a certain position the equilibrium becomes unstable and the electron proper to the large sphere makes a sudden jump into the smaller sphere. In the case in which the ratio of the radii of spheres is three to one, this instability occurs when the distance between the centres lies between 2·6 and 2·7 times the radius of the smaller sphere. After this has occurred both electrons remain inside the smaller sphere even if the larger sphere is taken away. We thus get a case which is analogous to the formation of a neutral compound molecule, and, as in the case of electrolytes, when this molecule is subsequently broken up one atom is positively and the other negatively charged.

We have seen that the important differences between the chemical elements are in all probability determined by the difference in the number of electrons in the atom rather than by the difference in size of the positive spheres. We have every reason to believe that the atoms of most of the elements contain a considerable number of electrons, so that there is no evident reason why they should behave in the same way as the extremely simply constituted atoms just now under discussion. It seems clear, however, on general grounds that if the atoms contain a large number of electrons they will attract one another whether they are like or unlike, and so will tend to coalesce into groups of more than one atom. That the forces between uncharged atoms will in general be attractive appears to follow from the fact that the electrons are more or less mobile. Under the influence of the electric field due to a neighbouring atom, these will arrange themselves so that their potential energy is diminished. It is to be remembered that although the atoms are electrically neutral there will be intense fields of force in their immediate neighbourhoods owing to the different geometrical distribution of the positive and negative electricity. The attraction between uncharged atoms is similar to that between an uncharged conductor and a charged sphere. This effect will take place whether the atoms are similar or dissimilar.

---

* Cf. Kelvin, *loc. cit.*

The combination of two atoms in this way will, in general, be accompanied by a transference of electrons from one to the other. The way in which this takes place is most conveniently described in terms of the idea of electronic pressure. Consider the atom which possesses N_2+1 electrons. Let us denote this by A and suppose it to be brought into immediate juxtaposition with the atom B, which contains N_3-1 electrons. We have seen that the electronic pressure of the atom A is very high while that of B is very low. In other words less work will be required to drag an electron out of A than out of B. Under these circumstances we should expect an electron to pass from A to B; so that in the compound thus formed A will carry one electronic unit of positive and B one electronic unit of negative electricity. After this transference has taken place A will contain N_2 electrons so that its electronic pressure will be very low. Thus no more electrons will pass over to B, especially since if another were to be transferred B would then have N_3+1 electrons, and this number corresponds to an atom with a very high electronic pressure. There is still another reason why no further transference should take place, and that is that the transference of the first electron produces an electric field between the atoms which tends to stop any further exchange.

If another atom of A were brought into the neighbourhood of the molecule AB it would not be able to transfer another electron to B, despite its own high electronic pressure. For if it did so B would then have N_3+1 electrons, an arrangement which is particularly likely to shoot off one electron. This action would certainly result in this case since it would be helped by the electric field from B to A. Thus A is clearly a monovalent electropositive element and B is a monovalent electronegative element. The molecule AB is a fully saturated molecule and will show no tendency to enter into further combination with any other elements.

Now consider the interaction between an atom C containing N_2+2 electrons and an atom D of another element containing N_3-2 electrons. The electronic pressure of C being higher than that of D, we should expect an electron to pass from C to D.

C will then have $N_2 + 1$ electrons and one unit of positive charge, whilst *D* will have $N_3 - 1$ electrons and one unit of negative charge. Now if *C* and *D* were uncharged their electronic pressures would be respectively higher and lower than before the transference. Thus there will be a tendency for another electron to pass from *C* to *D*, despite the opposition of the electric field established by the transference of the first electron. If this takes place there will be no further transference, since *C* and *D* now carry N_2 and N_3 electrons respectively; so that the transference of an additional electron would be opposed by the discontinuities in the electronic pressure as well as by the electric field produced by the two electrons already displaced from *C* to *D*. The compound *CD* is thus completely saturated. If *C* and *B* were placed in contact, an electron would evidently at once be displaced from *C* to *B*; but despite the increase in the electronic pressure of *C* thus caused no further transference could take place. For this would involve increasing the number of *B*'s electrons to $N_3 + 1$, an arrangement with a very high electronic pressure. Since this pressure would be assisted by an intense interatomic electric field, the two combined would effectually stop any tendency for a second electron to go to *B*. If, however, an uncharged atom of *B* were placed in the neighbourhood of *CB*, this restriction would not occur. This atom would have a low electronic pressure and would readily abstract an additional electron from *C*. Thus the compound molecule CB_2 would be formed. By similar reasoning to that which has preceded, this would be fully saturated and would have no tendency to combine with any other atoms. In a similar way the element *D* would form the fully saturated molecule DA_2 when allowed to combine with *A*. *C* and *D* are typical divalent elements, *C* being electropositive and *D* electronegative.

It is to be borne in mind that there is a difference between the energies liberated when *C* combines with *B* to form *CB* and when *CB* combines with *B* to form CB_2. The formation of CB_2 from *CB* is opposed to some extent by a preexisting electric field, whereas that of *CB* from *C* and *B* is not. In some cases this restraining influence may be sufficient to prevent the ready formation of the compound CB_2, and the action may stop at the *CB* stage. Nearly all the elements with high valencies furnish examples of an effect of this character. The intermediate stages are known as unsaturated compounds. A number of examples are given, selected from elements of different chemical families, in the accompanying table.

Element	Maximum valency	Compounds
Chlorine	+7	Cl_2O, Cl_3O_3, ClO_2, Cl_2O_7
	−1	HCl
Sulphur	+6	S_2Cl_2, SCl_2, SCl_4, SI_6
	−2	H_2S
Phosphorus	+5	PCl_3, PCl_5
	−3	H_3P
Carbon	+4	CS, CS_2; CO, CO_2
	−4	C_2H_2, C_2H_4, CH_4, C_2H_6
Iron	+3	Fe_2Cl_4, $FeCl_3$
	−5	
Copper	+2	Cu_2O, CuO
	−6	
Potassium	+1	KCl
	−7	

It is to be observed, as we have already seen is indicated by this theory in a general way, that the maximum valency of an element depends on whether it occurs in the compound as an electropositive or an electronegative element. Thus consider the element with $N_3 - 2$ electrons. We have seen that this can only take up two additional electrons before reaching the unstable stage. Two is therefore the maximum valency of the element when it occurs as the electronegative part of the compound. On the other hand it can part with $N_3 - 2 - N_2$ electrons before it reaches the other unstable configuration; so that this is the value of its maximum electropositive valency. A glance at the table will show that this property is very well illustrated by the compounds of chlorine, sulphur and phosphorus. It is obvious from these considerations that the terms electropositive and electronegative are merely relative. An element may be electropositive to one element and electronegative to another; in this case it will lose an electron when combined with the first element and gain one when combined with the second.

The function of the electrons in the elementary molecules such as H_2, O_2, N_2, Cl_2, Br_2, I_2, etc. is a matter of great interest. We have seen that similar atoms will show an attraction for one another, and they may be in equilibrium in pairs without the transference of an electron from one of the atoms to another. It is not however certain that such a state of equilibrium would be stable without the transference of one or more electrons. Suppose that the uncharged atoms contained N electrons, where N is a number lying between N_2 and N_3. We have seen that such an atom has a tendency both to emit, and to combine with, one or more electrons, the precise number depending on the value of N. It is conceivable that the arrangement $(N-1)(N+1)$ would be more stable than the arrangement NN. In the former case one atom would carry a positive and the other a negative charge, whilst in the latter case both atoms would be neutral. On the whole the evidence seems to be in favour of the elementary diatomic molecules containing oppositely charged as opposed to neutral atoms. The facts which bear on this question may be briefly summarized as follows:

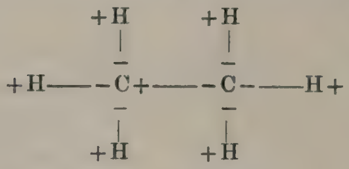
(1) Walden has found that in certain solvents the elementary solutes Br_2 and I_2 can be electrolysed and equal quantities of bromine or iodine are liberated at each electrode.

(2) In structural chemistry the bonds which combine like atoms are treated as being in every way similar to those which hold together unlike atoms.

(3) The inert gases, helium, neon, etc. which do not enter into combination with other elements have monatomic molecules. At the same time the metals have monatomic vapours for the most part and are averse to combination with one another, although they combine readily with electronegative elements.

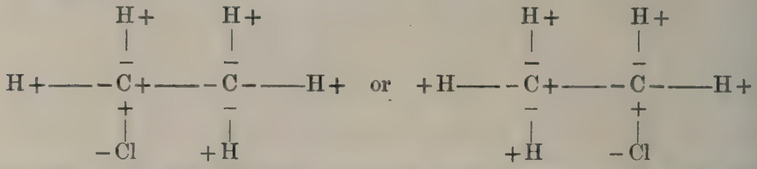
(4) The molecular refraction and dispersion of a substance might differ considerably according to whether the atoms occurred in oppositely charged pairs or not. There is no evidence of any well-marked difference between the optical properties of elementary substances, considered as a class, and those of compounds.

If we accept the view that the chemical bonds which unite different atoms of the same element involve the transference of an electron, it follows that the number of possible isomeric forms of many substances is greater than the number indicated by ordinary structural chemistry. Take for instance ethyl chloride C_2H_5Cl. This is obtained by substituting one atom of chlorine for one of the atoms of hydrogen in ethane C_2H_6. The structure of ethane would be represented according to the present hypothesis by



$$\begin{array}{ccccccc} & & +\mathrm{H} & & +\mathrm{H} & & \\ & & | & & | & & \\ & & - & & - & & \\ +\mathrm{H} & \text{——} & -\mathrm{C}+ & \text{——} & -\mathrm{C}- & \text{——} & \mathrm{H}+ \\ & & - & & - & & \\ & & | & & | & & \\ & & +\mathrm{H} & & +\mathrm{H} & & \end{array}$$

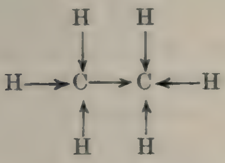
while that of ethyl chloride would be either



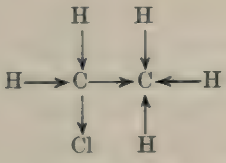
$$\begin{array}{ccccccc} & & \mathrm{H}+ & & \mathrm{H}+ & & \\ & & | & & | & & \\ & & - & & - & & \\ \mathrm{H}+ & \text{——} & -\mathrm{C}+ & \text{——} & -\mathrm{C}- & \text{——} & \mathrm{H}+ \\ & & + & & - & & \\ & & | & & | & & \\ & & -\mathrm{Cl} & & +\mathrm{H} & & \end{array} \quad \text{or} \quad \begin{array}{ccccccc} & & \mathrm{H}+ & & \mathrm{H}+ & & \\ & & | & & | & & \\ & & - & & - & & \\ +\mathrm{H} & \text{——} & -\mathrm{C}+ & \text{——} & -\mathrm{C}- & \text{——} & \mathrm{H}+ \\ & & - & & + & & \\ & & | & & | & & \\ & & +\mathrm{H} & & -\mathrm{Cl} & & \end{array}$$

according as the chlorine entered into combination with the carbon atom containing a positive unit of electric charge or not. It is obvious that ordinary structural chemistry does not admit of any dissymmetry of this kind, and its position in this respect is supported by the facts. Although attempts have been made to prepare such compounds as ethyl chloride in different ways, so as to isolate different isomeric forms if they existed, they have invariably been unsuccessful. The ordinary structural formulae appear to be quite capable of taking account of all the different isomeric forms of carbon compound which can be prepared. This, however, is no very conclusive evidence against the view above as to the nature of the bonds, for in most cases it is clear that one of the two possible forms would be much more stable than the other; so that even if the less stable form were produced at first, it would immediately be changed into the other. We should expect the number of structurally possible compounds always to be greater than the number of those which can be actually isolated. It is clear that the number of these possible electrical isomers will increase rapidly with the number of carbon atoms in the compound.

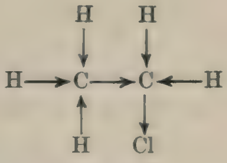
This method of looking at chemical combination gives a definite physical meaning to the bonds of structural chemistry. They represent the directions of the electric fields between the atoms. The bonds are to be regarded as starting on a negative electron and ending on a positive or negative charge. A bond extending between two points may therefore be either positive or negative. This quality is not taken account of in ordinary structural chemistry. The position of the end of the bond where the negative charge is may conveniently be indicated by an arrow pointing towards it. This enables us to omit the cumbersome positive and negative signs in the formulae on p. 576. These would then become



$$\begin{array}{ccccc} & H & & H & \\ & \downarrow & & \downarrow & \\ H\rightarrow & C & \rightarrow & C & \leftarrow H \\ & \uparrow & & \uparrow & \\ & H & & H & \end{array}$$



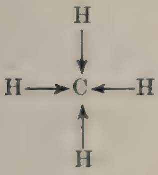
$$\begin{array}{ccccc} & H & & H & \\ & \downarrow & & \downarrow & \\ H\rightarrow & C & \rightarrow & C & \leftarrow H \\ & \downarrow & & \uparrow & \\ & Cl & & H & \end{array}$$



$$\begin{array}{ccccc} & H & & H & \\ & \downarrow & & \downarrow & \\ H\rightarrow & C & \rightarrow & C & \leftarrow H \\ & \uparrow & & \downarrow & \\ & H & & Cl & \end{array}$$

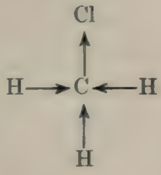
The application of this method of drawing structural formulae may also be illustrated by the successive chlorine substituted methanes, as follows:

1



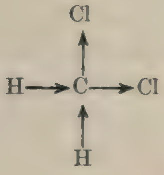
$$\begin{array}{ccc} & H & \\ & \downarrow & \\ H\rightarrow & C & \leftarrow H \\ & \uparrow & \\ & H & \end{array}$$

2



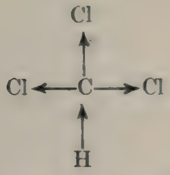
$$\begin{array}{ccc} & Cl & \\ & \uparrow & \\ H\rightarrow & C & \leftarrow H \\ & \uparrow & \\ & H & \end{array}$$

3



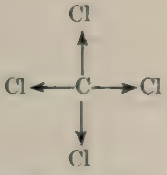
$$\begin{array}{ccc} & Cl & \\ & \uparrow & \\ H\rightarrow & C & \rightarrow Cl \\ & \uparrow & \\ & H & \end{array}$$

4



$$\begin{array}{ccc} & Cl & \\ & \uparrow & \\ Cl\leftarrow & C & \rightarrow Cl \\ & \uparrow & \\ & H & \end{array}$$

5



$$\begin{array}{ccc} & Cl & \\ & \uparrow & \\ Cl\leftarrow & C & \rightarrow Cl \\ & \downarrow & \\ & Cl & \end{array}$$

It is clear that the total positive charge on an atom in electronic units is obtained by subtracting the number of bonds which point towards the atom from the number which point away from it. Thus the carbon atom in each of the successive compounds 1 to 5 will be charged with – 4, – 2, 0, + 2, + 4 electronic units of electricity respectively. The formulae have been drawn, of course, on the supposition that each hydrogen atom carries one unit of positive charge and each chlorine atom one unit of negative charge.

Even when all the valency electrons existing in a compound molecule are saturated, there will still be a considerable external field, just as a neutral electric doublet gives rise to an external field. These forces would not be so intense as those which come into play during chemical combination proper, but they are probably capable of accounting for such phenomena as hydration, cohesion, surface tension, latent heat and the properties arising from molecular forces generally.

In considering the nature of the bonds connecting atoms of the same element it is interesting to observe that the phenomenon of self-combination is only shown to any considerable degree by two elements, carbon and silicon, and that both these elements lie exactly in the centre of their respective series in the periodic table. Their electrochemical properties are therefore neither markedly electropositive nor electronegative, and they ought to show an almost equal tendency to enter into combination in either sense. This is just the kind of condition of the atom that we should expect would give rise to self-combination, on the view that we have adopted as to the mechanism of this action. For such an atom would be almost equally stable whether it formed the positive or the negative end of a chemical bond. It would show a greater degree of adaptability to the effect of the remaining groups with which the atoms were combined than would atoms which were not so constituted.

If this view of chemical combination is the correct one, it becomes a matter of great importance to determine the sign and magnitude of the charge carried by each atom in different compounds. The most generally applicable method is, of course, the electrolytic one. In the case where one atom is deposited at the cathode it presumably carries a positive charge in the compound; since, so far as the writer is aware, no cases are known where the solvent is capable of reversing the polarity of the ions formed by a given electrolyte.

The phenomenon of magnetism also seems to be capable of shedding some light on this phenomenon. Townsend has shown that the magnetic permeability of solutions of salts of iron containing the same amount of iron has the same value for all ferric salts; it also has the same value, but one which is different from the preceding, for all ferrous salts, whilst for the ferricyanides it has uniformly the value zero, i.e. the ferricyanides are non-magnetic. These results show that the magnetism of iron compounds is an atomic property but indicate that it depends on the electric charge carried by the iron atom. From chemical reasons we believe that the iron atom in ferric salts carries three units of positive electricity, whereas in the ferrous salts it contains only two and in the ferricyanides it occurs in the electronegative part of the molecule. Thus by making experiments on the magnetic properties of any iron salt we could tell what the electric charge in the iron atom in the molecule was.

Thomson suggests that considerations of this kind may account for the magnetic properties of elementary oxygen. It is well known that both oxygen and ozone are strongly paramagnetic. On the view of atomic combination above, the oxygen molecule may be expected to contain one atom which is positively charged. Now in all the known compounds of oxygen it probably functions as the electronegative constituent. Even when combined with the strongly electronegative element chlorine, oxygen appears to be electronegative. This is indicated by the high valency of chlorine in the higher oxides ClO_2 and Cl_2O_7, and by the oxide Cl_2O being to some extent acid forming. Thus elementary oxygen is the only form in which an electropositive oxygen atom appears to exist, and this may explain why oxygen and ozone are the only substances in which oxygen is paramagnetic.

Attention has recently been called by K. G. Falk and J. M. Nelson* to a number of facts in structural chemistry which seem to support Thomson's view of directed valencies. Among these the following may be mentioned:

(1) A hydrocarbon which is thought to have the constitution $[C(C_6H_5)_3]_2$ conducts the electric current when in solution. As the electric current does not appear to produce permanent chemical change the ions have been thought to be $-C(C_6H_5)_3$ and $+C(C_6H_5)_3$.

(2) The symmetrical saturated dicarboxylic acids of the type $CO_2H(CH_2)_p-CH_2-(CH_2)_pCO_2H$ show definite differences in their physical properties when they are compared as a group with the rather similar acids $CO_2H(CH_2)_p-(CH_2)_pCO_2H$. In the former case the directed valencies may be arranged symmetrically whereas in the latter they cannot.

(3) The quantitative yields of the various isomers formed when the unsymmetrical hydrocarbons of the ethylene series

---

* *Proc. Amer. Chem. Soc.* vol. XXXII. p. 1637 (1910).

combine with the haloid acids are readily interpreted on this view, as also are the chemical properties of the diazo-compounds and a number of other substances.

It is necessary to add that most of these facts have been accounted for by chemists in other ways which seem fairly satisfactory.

Reviewing the whole question broadly it seems quite likely that in a great number of cases of chemical combination transference of electrons between the atoms will not occur. There seems to be little doubt that the forces between uncharged atoms can be sufficiently great to account for the energy of chemical combination in a great number of cases. Definiteness of valency can also be accounted for in this way. For instance, if the neutral carbon atom possesses four electrons arranged at the corners of a regular tetrahedron, the directions of maximum electric intensity in the field of the neutral atom will be along lines possessing a similar tetrahedral symmetry. The safest course to adopt at present would appear to be that of restricting the interpretation of valency bonds as representing electronic transference to those cases only in which the possibility of electrolytic dissociation has been demonstrated.

The Structure of the Positive Electricity.

The foregoing, necessarily brief, review of chemical phenomena shows that there is a very close correspondence between the properties of the elements and those required by the atoms considered by Thomson. It is not likely that the hypothesis of a sphere of positive electrification of a uniform volume density is essential in order to arrive at conclusions of the same general character as those which have been indicated. It is probable that somewhat similar conclusions about questions of stability would hold if the volume density of the positive electrification, instead of being uniform, were greatest at the centre of the sphere. On the other hand, we have seen that it is impossible to construct a definite atom out of indefinitely small elements of positive and negative electrification acting on each other according to the classical laws of electrodynamics. It is necessary to introduce something else in order to account for the actual size of the atoms.

The sphere of continuously distributed positive electrification has the merit of lending itself readily to calculation and, as we have seen, it gives a satisfactory account of many of the properties of the chemical elements. On the other hand there are some rather striking properties which it leaves unexplained, at any rate in its simplest form. Consider for example the phenomenon of radioactive transformation. The chemical and spectroscopic properties of the typical radioactive elements radium, thorium and uranium are not sharply different from those of the elements which do not exhibit radioactivity; so that it does not seem likely that they have a constitution radically different from that of the others. But we know that the atoms of the radioactive elements are continually emitting atoms of helium and turning into other elements of lower atomic weight. Thus an emission of part of the positive sphere itself is a possibility which has to be contemplated. To account for these phenomena Thomson has suggested that the atoms of the elements of higher atomic weight are made up of combinations of sub-spheres like that which may be supposed to constitute the helium atom; but on such a view it is not easy to see why the atoms should possess a different order of stability from that of their compounds, without importing something else into the theory. Another point which deserves some consideration is the fact that there is an upper limit to the weights of atoms. All the elements of higher atomic weight than bismuth (208) are radioactive and therefore unstable. The striking feature of this instability is that it appears to affect the positively charged part of the atom, as well as, if not more than, the negatively charged constituents.

As it stands, the positive sphere gives no account of the mass of the atom. Unless all the methods of estimating the number of electrons in the atoms are entirely misleading, a contingency which is very unlikely in view of the excellent agreement given by entirely different methods, practically the whole of the mass of matter must belong to the positively electrified parts of it. Now the electromagnetic inertia of the positive spheres of Kelvin and Thomson is negligible compared with that of a single electron, so that the greater part of the mass is entirely unaccounted for by this theory.

The following hypothesis appears to offer a possibility of explaining these facts and at the same time of retaining the important features of the positive sphere. Suppose that the positive electricity, instead of being uniformly distributed, is in the form of electrons whose charge is E and mass M, where E is very small compared with the numerical value of the charge e of the negative electrons. Since the electromagnetic mass of such particles is proportional to E^2/R, where R^3 is comparable with their volume, it is clear that if R were small enough the positive electricity could be made to carry most of the inertia of the atom. If $e = \nu E$, then $r/R = \nu e/m \div E/M$, where the small letters refer to the negative electrons. If the hydrogen atom contains only one negative electron, then E/M is the value of the corresponding quantity for the hydrogen atom in electrolysis. r would be greater than R in the proportion of something comparable with 1000ν, so that the positive electron would have to be confined to a much smaller space even than the negative. This would lead to the difficulty about the definiteness of the atom already alluded to, if the law of force between these positive electrons were that of the inverse square. Let us suppose that at very small distances this law does not hold but is replaced by something more complicated, let us say

$$+\frac{a}{r^2}-\frac{b}{r^{p_1}}+\frac{c}{r^{p_2}},$$

where $p_2 > p_1 > 2$. At very small distances the third term would give a repulsion and keep the positive electrons from joining together, and the first term would give the usual law of force at large distances. The middle term would cause the positive electrons to attract one another at certain distances. This would make them aggregate into clusters which, if the constants were of suitable magnitude, would be of dimensions comparable with that of the atoms. The positive electrons would be regularly distributed inside so that such clusters would behave very much like a continuous distribution of positive electrification, provided E were sufficiently small.

On this hypothesis there would be a definite positive atom capable of existing without the presence of a negative electron.

Such atoms would be able to coalesce if the a/r^2 term were neutralized by the presence of a negative electron, but not otherwise. In this way more complicated atoms could build up, and the conditions for the equilibrium of the electrons in them would be similar to those of Thomson's theory. There would be a limit to the size of atoms which could be built up in this way, because with the larger atoms it would be difficult to arrange the negative electrons so as to prevent a whole unit of positive electricity (νE) from ever becoming unstable.

It will be urged that there is no experimental evidence in favour of the existence of particles carrying a charge less than e. That is true, but if E/e is very small it is questionable whether any experiments which have been made would be capable of detecting their existence. Their mutual attractions would prevent an electron from getting away from an atom with any considerable number of them. Another objection is that all known chemical atoms are either neutral or carry a charge which is an integral multiple of $\pm e$. It may be that the elementary positive atom also has a charge which is equal to an integral multiple of e, or that the law of force somehow makes the stable systems neutral. It must at least be admitted that these suggestions are not more artificial than the hypothesis that the atoms are provided with a sphere of positive electrification just sufficient to neutralize the electrons present.

Collisions between systems of this kind would be rather different from those between atoms made up of positive electricity of uniform density, and might be expected to give rise to a relatively high percentage of sharply deflected α rays, such as was found by Rutherford and Geiger.

We shall not pursue this subject further. The deflexions of the α rays through large angles and the scattering of X rays by light atoms, which were considered in Chapter XIX, agree much better with Rutherford's view that the positive electricity in the atom is concentrated in a minute region of it than with the uniform sphere of positive electrification. This position is made stronger still by the considerations brought forward in the two next sections. It will be noticed that if the linear dimensions of the nucleus are small enough the whole mass of the atom may be of electromagnetic origin.

The Radioactive Elements and the Periodic Law.

The study of the chemical properties of the radioactive elements* has brought to light a number of facts of the highest importance which bear on the relation between the chemical properties of the elements and their atomic weights. Soddy pointed out that when a radioactive element A was converted into a second element B with accompanying emission of an α ray, then the chemical properties of B would be those of an element in the next column but one before A in the periodic table. Russell† showed that if a β ray were emitted instead of an α ray the new element would be found in the next column beyond that of the first element. Thus, to illustrate the case by considering one particular chemical property, the emission of an α ray diminishes the electropositive valency by two, whereas the emission of a β ray increases it by one. In the case of the thorium series for example, if we neglect branch products, the successive changes are exhibited in the following table:

$$\underset{\text{IV}}{\text{Th}} \xrightarrow{\alpha} \underset{\text{II}}{\text{Msth I}} \xrightarrow{\text{Rayless}} \underset{\text{III}}{\text{Msth II}} \xrightarrow{\beta} \underset{\text{IV}}{\text{Ra Th}} \xrightarrow{\alpha} \underset{\text{II}}{\text{Th X}} \xrightarrow{\alpha} \underset{\text{O}}{\text{Em}} \xrightarrow{\alpha} \underset{\text{VI}}{\text{A}} \xrightarrow{\alpha} \underset{\text{IV}}{\text{B}} \xrightarrow{\beta} \underset{\text{V}}{\text{C}} \rightarrow$$

The numerals underneath represent the number of the column in the periodic table (or the value of the electropositive valency). The character of the change is indicated above the arrows.

But the matter goes further than this. In some cases an α ray change is followed by two successive changes in which an α ray is not expelled. The last element then occupies the same column in the periodic table as the original element. In such cases these two elements have, so far as can be ascertained, identical chemical properties, and are incapable of being separated by chemical methods. There is evidence‡ also that their emission spectra are identical. These results are surprising at first sight since the atomic weights of the elements in question must differ by the atomic weight of helium approximately, or about four units. In contrast to their chemical properties, the radioactive properties of two such elements are quite different. These phenomena receive a plausible explanation on the view, advocated by Rutherford, that the atoms are built up of electrons revolving round a massive central nucleus of small dimensions, if we adopt the hypothesis that both the α and β particles are ejected from the nucleus itself. For the chemical and spectroscopic properties of such an atom will be determined almost entirely by the charge of the nucleus and hardly at all by its mass. The ejection of an α particle, which is a positively charged helium atom with twice the electronic charge, will diminish the total positive valency of the nucleus by two, and the ejection of an electron will increase it by one. The resulting nuclei will be those appropriate to atoms whose positions in the periodic table are those actually found.

---

* F. Soddy, *Chem. News*, vol. CVII. p. 97 (1913); G. v. Hevesy, *Phys. Zeits.* vol. XIV. p. 49 (1913); K. Fajans, *Phys. Zeits.* vol. XIV. pp. 131, 136 (1913); *Verh. d. Deutsch. Phys. Ges.* vol. XV. p. 240 (1913).

† *Chem. News*, vol. CVII. p. 49 (1913).

‡ Russell and Rossi, *Roy. Soc. Proc.* A, vol. LXXXVII. p. 478 (1912).

This position is strengthened by a number of other lines of evidence. Thus J. J. Thomson* has discovered that neon consists of two gaseous elements having the same spectrum and chemical properties but differing in atomic weight. We have seen also that the X-ray spectra of the elements are not accurately functions of the atomic weight but are accurately determined by successive whole numbers. The sequence of these numbers is that of the atoms in order of ascending atomic weight except for the elements whose positions in the periodic table are anomalous. It seems natural to interpret the "atomic numbers" as measures of the charge carried by the positively charged nucleus. These facts do not receive any obvious explanation on the Kelvin-Thomson theory of the atom.

Bohr's Theory of the Behaviour of Electrons in Atoms.

This theory†, which frankly discards dynamical principles, has already been referred to in Chap. XVI. As its chief success so far has been in the explanation of the frequency of spectral lines and other phenomena which depend intimately on atomic structure, it may be permissible to recapitulate here the chief assumptions on which it rests. It is based on the idea of negative electrons revolving round a massive positive nucleus of negligibly small dimensions, the mutual force following the electrostatic law of the inverse square. Stability and definiteness of frequency, which, as we have seen, are dynamically impossible with such a structure, are secured by the introduction of the two following hypotheses:

---

* *Roy. Soc. Proc.* A, vol. LXXXIX. p. 1 (1913).

† N. Bohr, *Phil. Mag.* vol. XXVI. pp. 1, 476, 857 (1913), vol. XXVII. p. 506 (1914).

(1) That the stable motions are those for which the angular momentum of the electrons is an integral multiple of $h/2\pi$, and

(2) That radiation is emitted when there is a change from one possible configuration of the electrons to another. The radiation thus emitted is assumed to be monochromatic and its frequency ν to be determined by the equation $W_1 - W_2 = h\nu$, where h is Planck's constant and W_1 and W_2 are the energies of the system in the two configurations under consideration.

These assumptions lead immediately to the following formula for the possible frequencies which may be emitted by the simplest atom (that which contains a single electron when neutral):

$$\nu = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{\tau_2^2} - \frac{1}{\tau_1^2}\right) \quad \text{.................}(29),$$

where τ_2 and τ_1 are whole numbers. When $\tau_2 = 2$ and τ_1 takes successive integral values this gives Balmer's series. The strong point in favour of this theory is that it determines the absolute constant in the spectral series in terms of the fundamental constants m, e and h, and the agreement is exact to the accuracy within which these constants are known. The spectroscopic value of the constant K in Balmer's series is $3{\cdot}290 \times 10^{15}$. When the best values of e, m and h are used the calculated value of the factor $\frac{2\pi^2 me^4}{h^3}$ in (29) is $3{\cdot}26 \times 10^{15}$. By putting $\tau_2 = 3$ another known hydrogen series is obtained.

When the central charge has twice the value just considered and there is a single electron revolving round it, we should expect to get the spectrum of a helium atom which had lost both of its electrons (an α particle) and subsequently combined with a single electron. As a matter of fact the series calculated for this system include Pickering's* series and certain series recently observed by Fowler†. None of these series have been obtained in the laboratory except in the presence of helium.

---

* *Astrophys. Journ.* vol. IV. p. 369 (1896), vol. v. p. 92 (1897).

† *Month. Not. Roy. Astr. Soc.* vol. LXXIII. (1912).

Another consequence of these assumptions is that the constant N_0 in Rydberg's formula for spectral series should have the same value for every element. The theory also gives the correct order of magnitude for the dimensions of atoms and for the work necessary to liberate ions from atoms by collisions. It is also in agreement with Whiddington's law connecting the atomic weights of the elements with the velocity of cathode rays which will excite the emission of characteristic X-rays from them.

In the last paper (*Phil. Mag.* vol. XXVII. p. 506) Bohr has modified this position somewhat. The energies in the different configurations (stationary states) are treated as empirically given by the laws of spectral series and the value of the constant K is determined from the condition that for very slow vibrations, corresponding to large values of τ, the frequencies must agree with those calculated from the classical electrodynamics. The condition (1) above then appears as a particular case restricted to circular orbits; in general it is replaced by $T_n = \frac{1}{2} n h \omega_n$, where T_n is the mean kinetic energy of the electrons, and ω_n the frequency, in the nth stationary state.

Whatever view may be taken as to the character of the underlying assumptions there is no doubt that this theory has been much more successful in accounting quantitatively for the numerical relationships between the frequencies of spectral lines than any other method of attack which has yet been tried. Although the assumptions conflict with dynamical ideas they are of a very simple and elementary character. The fact that they conflict with dynamics does not appear to be a valid objection to them, as there are a number of other phenomena, the temperature radiation for example, which show that dynamics is inadequate as a basis for a complete explanation of atomic behaviour.

CHAPTER XXII

GRAVITATION

General Characteristics.

The position of physical theories of gravitation to-day is almost as speculative as it has been ever since the days of Newton. Nevertheless we cannot afford entirely to overlook the bearing of the electron theory on gravitational action, particularly as we have found the theory to be capable of giving a fair account of the other known physical phenomena. Paradoxical though it may seem, it is quite likely that one of the chief difficulties in the way of a physical theory of gravitation lies in the extreme simplicity of the known laws of gravitational action. The prospect of addition to our knowledge of gravitation by experiment is not very hopeful on account of the smallness of the forces concerned. These are large enough, of course, when we are dealing with the enormous aggregations of matter familiar to astronomy, but they are exceedingly small with masses which can be controlled in the laboratory.

The fundamental law of gravitation, which was enunciated by Newton, may be expressed by means of the equation

$$F = k\frac{dm\,dm'}{r^2},$$

where F is the attractive force between two material particles whose masses are dm and dm' at a distance r apart, and k is the gravitational constant which is equal to $6{\cdot}6 \times 10^{-8}$ in C.G.S. units. This inverse square law of attraction is the foundation of the astronomical calculations of the orbits of the heavenly bodies. As these are very exact it follows that the index of r is 2 with great precision.

The value of k is also very accurately independent of the *nature* of the attracting materials. This is proved by experiments made with pendulums constructed from different substances. These show that the weight of a body, at any point of the earth's surface, divided by its mass is a constant independent of the nature of the body. This law of proportionality between weight and mass is one of the most accurately known of the experimental laws of physics.

The inverse square law at once suggests a relation with electrostatic forces. If neutral matter consists of oppositely charged elements it is clear that two uncharged particles will attract one another if the attraction between unlike elements slightly exceeds the repulsion between like elements. There is, however, a well-known difference between gravitational and electrical effects which calls for consideration. The space inside a closed electrical conductor is found to be completely shielded from the effects of an external field of electrical force, whereas an electrical conductor exerts no such shielding effect from the action of gravitational force. Nevertheless this is not a valid objection to the view that the two forces are of the same nature. It is qualitatively obvious that the electrostatic shielding arises from the opposite displacement of the two kinds of electricity along the conductor in such a way as to tend to annul the field inside. A precise calculation shows that the compensation is exact if the law of electrostatic force is the inverse square. In the case of gravitational action the effect of a conducting screen will be to increase the attraction between two material bodies A and B rather than to diminish it. For, let the screen C surround B as in the figure. The fundamental property of a conductor lies in

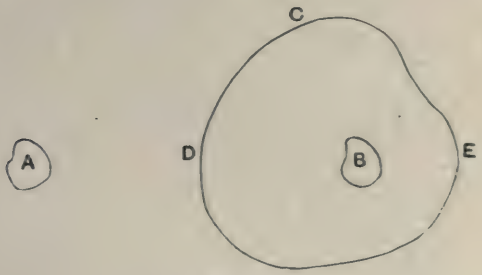


Fig. 57.

the mobility of some of its ultimate electrified parts. Let some of these (P) be attracted by neutral matter and others (Q) repelled. The effect of A on the screen will be to make the parts P move towards D and the parts Q towards E. Thus the force urging B towards D is increased by the presence of the screen C, owing to the formation of an attractive layer at D and a repulsive layer at E. It is evident that this additional force will be exceedingly minute. In a similar way gravitating matter might give rise to a field of electric force inside a hollow conductor, but with any probable hypothesis about the action of gravity on electrons the resulting electric forces are too small to detect.

Comparison with Electrostatic Forces.

Before proceeding further it is desirable to emphasize the smallness of gravitational attraction compared with the forces between the electrons composing the attracting matter. We may illustrate this by comparing the gravitational attraction between two material particles of masses m_1 and m_2 with the electrical attraction between two particles of like masses, of which the first (m_1) consists solely of positive electrons and the second (m_2) consists solely of negative electrons. The magnitude of the gravitational attraction is $6{\cdot}6 \times 10^{-8}\, m_1 m_2/r^2$ dynes. Let N_1 be the number of positive electrons in m_1 and n_2 the number of negative electrons in m_2. Then if E and e are the respective charges on the electrons, we have

$$\frac{n_2 e}{m_2} = \frac{e}{m} \text{ for the negative electrons} = 5{\cdot}4 \times 10^{17} \text{ E.S. units.}$$

Let us take, as a lower limit for E/M for the positive electron, the value of this quantity for the hydrogen atom in electrolysis. We then have

$$\frac{N_1 E}{m_1} = \frac{E}{M} = 3 \times 10^{14} \text{ electrostatic units.}$$

So that the gravitational attraction

$$= \frac{6{\cdot}6 \times 10^{-8}}{5{\cdot}4 \times 3 \times 10^{31}} \frac{N_1 E \times n_2 e}{r^2} \text{ ....} (1)$$

$$= 4{\cdot}1 \times 10^{-40} \text{ times the electrostatic attraction.}$$

Looked at from this point of view it is rather surprising that the electrostatic forces do come so near to balancing one another and it is quite clear that it will only be necessary to make a minute change in the law of force between electrons in order to modify it sufficiently to take account of gravitational attraction. The change which we shall have to make will be quite beyond anything which could be detected by direct experiments on the forces exerted between electric charges.

Law of Force between Electrons.

Perhaps the simplest assumption that we can make in order to account for gravitational attraction is that the repulsive force between two electrons, whose charges are E_1, E_2 respectively, is not accurately equal to $\frac{1}{4\pi}\frac{E_1 E_2}{r^2}$, as we supposed in Chapter I. It may be that the force is not determined entirely by the charges of the electrons but depends to a slight extent on the geometrical configuration and the state of motion of the charges as well. In order to take this into account we may write the force between E_1 and E_2 in the form $\frac{1}{4\pi}\frac{E_1 E_2}{r^2}F(E_1 E_2)$, where F denotes a function of the configuration and state of motion of E_1 and E_2. The magnitude of F can only differ very slightly from unity. Let us consider two material particles P and Q at a distance r from each other and such that P consists of N_1 positive electrons of charge E_1 and n_1 negative electrons of charge e_1 and the corresponding magnitudes for Q are indicated by the suffix 2. It is evident that the total repulsion between P and Q is

$$[N_1 N_2 E_1 E_2 F(E_1 E_2) + N_1 n_2 E_1 e_2 F(E_1 e_2) + N_2 n_1 E_2 e_1 F(E_2 e_1) \\ + n_1 n_2 e_1 e_2 F(e_1 e_2)]/4\pi r^2.$$

If the two particles are uncharged, $N_1 E_1 = -n_1 e_1$ and $N_2 E_2 = -n_2 e_2$. Let us also suppose that the electrons of like sign are identical in both bodies and that all the electrons of whatever sign have equal charges. Then

$$E = E_1 = E_2 = -e_1 = -e_2 = -e$$

and the repulsion becomes

$$\frac{N_1 N_2 E^2}{4\pi r^2} \times \{F(E_1 E_2) + F(e_1 e_2) - F(E_1 e_2) - F(E_2 e_1)\} \ldots\ldots (2).$$

The simplest assumption that we can make is that the electrons of like sign are in the same condition in whatever substance they occur. In this case the functions F will all be constants characteristic of the electrons. It also follows by symmetry that $F(E_2e_1) \equiv F(E_1e_2)$. Since all the functions are very nearly equal to unity, we may write

$$F(E_1E_2) = 1 + a, \quad F(e_1e_2) = 1 + b, \quad F(E_1e_2) = F(E_2e_1) = 1 + c,$$

where a, b and c are very small constant quantities. The repulsion between two neutral particles is therefore equal, on this supposition, to $\frac{N_1N_2E^2}{4\pi r^2}(a + b - 2c)$. This will be negative if $2c$ is greater than $a + b$, so that in order to account for gravitational attraction it is necessary to suppose that the attraction between two unlike charges is slightly greater than the repulsion between two like charges.

On the further hypothesis that all the mass of the substances is electromagnetic it will be proportional to the numbers N_1 and N_2 of the electrons which the particles contain, except in so far as the electromagnetic mass of an electron may vary somewhat in different atoms. If M_1 and M_2 are the masses of the particles at P and Q respectively and if M is the mass of a positive and m that of a negative electron, then $N_1 = \frac{M_1}{M + m}$ and $N_2 = \frac{M_2}{M + m}$, so that the repulsion becomes

$$\frac{M_1M_2}{4\pi r^2}\left\{\frac{E^2}{(M + m)^2}(a + b - 2c)\right\} \quad \ldots\ldots\ldots\ldots\ldots(3).$$

Since the quantity in brackets is a universal constant, subject to the hypotheses which we have made, the attraction accords with the Newtonian law.

The preceding result is based on the hypothesis that the mass of the electrons is not appreciably different in different substances. We shall see in the next paragraph that we cannot be sure that this is the case.

The Atomic Weights of the Elements.

It is now fairly certain that the number of negative electrons in any atom is equal to about half the atomic weight or, at all events, is comparable with the atomic weight. As the value of e/m for the negative electrons is very large it follows that only a very small proportion of the mass of an atom can arise from the electromagnetic mass of the negative electrons. If the mass of the atoms is of electromagnetic origin it must be associated almost entirely with the positive electricity. As we have seen, there is a considerable body of evidence in favour of the view that the positive electricity is confined to a very minute region of the atom, so that practically the whole of the atomic mass will reside in this nucleus. The estimated linear dimensions of the nucleus are very small (about the same as the linear dimensions of the negative electron) so that considerable overlapping of the fields of the constituent positive electrons may be expected. We should therefore expect the electromagnetic mass of such a nucleus not to be equal to the sum of the masses of the constituent positive electrons on account of the mutual interference of their electric fields. It is impossible to say precisely what difference would be expected to arise in this way, but it is likely to be only a small fraction of the total mass.

The expulsion of α rays from the atoms of the radioactive elements shows that the nucleus is made up of separable parts. However, the α particles are atoms of helium and have very nearly four times the mass of the lightest known positively charged particle, the positive hydrogen ion. Thus unless we accept the hypothesis that there is more than one kind of positive electron we cannot regard the α particle as such an ultimate unit. If we take the hydrogen ion to be the positive electron, the value of e/m which it possesses shows that its linear dimensions are about 10^{-4} of that of the nuclei. If such positive electrons were distributed fairly uniformly throughout the volume of the nucleus the overlapping of the fields would have little influence on the electromagnetic mass. The electromagnetic mass of an atom would then be very close to the sum of the masses of the positive electrons, and the truth of the statement at the end of the preceding paragraph would follow. We cannot, however, be sure of this deduction as we are completely ignorant of the constitution of the hypothetical nucleus.

It might be thought that further information might be obtainable from a consideration of the atomic weights of the elements. If the mass of an atom can be treated with sufficient accuracy as the sum of the masses of the constituent positive and negative electrons and if each kind of atom can be regarded simply as a neutral combination of a given number of such units then, to the same degree of accuracy, the atomic weights of the different elements will be proportional to whole numbers. This is on the supposition that there is only one kind of positive and one kind of negative electron. If there are n different kinds of electrons of masses m_1, m_2, etc., then the masses of the atoms will be of the form $a_1 m_1 + a_2 m_2 + \dots$, where a_1, a_2, etc. are integral numbers starting from zero. It is well known that the atomic weights of the elements exhibit an approximation to whole numbers which cannot be fortuitous*; but the deviations from integrals, or from the linear law just referred to, cannot now be regarded as likely to arise from modifications of the electromagnetic mass of the individual electrons. The work of Soddy and others (see p. 583) on the chemical properties of the radioactive elements has shown that elementary bodies having identical chemical properties and different atomic weights frequently coexist. It is therefore likely that a great many so-called chemical elements are mixtures of atoms having the same chemical properties but differing in mass. The experimentally determined atomic weight is, on such a view, the average of the weights of the constituent atoms and depends on the relative proportion in which they are present.

There is another point which deserves consideration in this connection. The two lightest positively charged particles with which we are familiar are the hydrogen ion and the α particle. According to the views of Rutherford and Bohr both of these particles consist only of positive electricity. The α particle has twice the actual charge and approximately half the specific charge (e/m) of the hydrogen ion. It is quite possible that the α particle is made up of four hydrogen ions cemented into a nucleus by means of two negative electrons, but it has not, so far, been shown that the α particle can be broken up into anything smaller; so that we shall examine tentatively the consequences of supposing that it is a fundamental element of positive electricity independent of the hydrogen ion. In that case we have to deal with two positive electrons, one of which has associated with a given charge approximately twice as much mass as the other. On p. 591 we considered gravitational attraction from the point of view that it arose from a minute lack of compensation between the attractions and repulsions of the ultimate electric charges. By supposing the additional forces to be determined by the electric charges we arrived at equation (3) in agreement with the Newtonian law of gravitational attraction. This was on the hypothesis that the value of e/m for all the electrons of given sign was the same. If some of the positive electrons, let us say the α particles, have a different value of e/m from the others, hydrogen ions, the same argument will lead to the conclusion that the weights of different elements will not be proportional to their masses if their respective atoms contain hydrogen ions and α particles in different proportions. Thus if we are to retain the law of proportionality between mass and weight, which has always been found to hold with high accuracy, it is necessary to suppose that the modification of the law of force does not depend simply on the charge of the electrons. It must depend quite directly on the mass which is associated with a given ultimate element of electric charge. The necessary change in the law of force will be considered in the next section. It is to be remembered that the necessity for considering this question depends entirely on the hypothesis that there are different positive electrons having specific values of e/m. There is no *logical* necessity for any such hypothesis since, as we have seen, it is quite possible that the nucleus of the helium atom may prove to consist of four hydrogen ions held together by two negative electrons.

---

* Strutt, *Phil. Mag.* [6] vol. I. p. 311 (1901).

Forces between Charges Modified by Mass.

We shall now consider the consequences of the hypothesis that the force between two ultimate elements of electric charge is not determined solely by the quantity of electricity present in the elements but depends to a slight extent on the mass associated therewith as well. The results of this hypothesis are free from contradiction so far as the facts which have been reviewed are concerned. Such a hypothesis seems a natural one from another standpoint. We might regard the ordinary form of the electrostatic law of force, which makes the force between neutral systems vanish, as an ideal which is attained when the ultimate elements of charge are devoid of relative motion and have an infinitesimal volume density. According to the electron theory these conditions are not realized in any actual material system. Another way of describing very much the same position is to say that the electrostatic law of force represents what would occur if the infinitesimal elements of charge were at rest in the undisturbed aether, that is to say in aether devoid of energy. The classic electrostatic law of force would then represent exactly what would occur if the aether were homogeneous. But in the immediate neighbourhood of an actual element of electric charge there is an immense disturbance going on and, as we have seen, the mass of the element is a measure of the energy of this disturbance. We might therefore expect that the field of force emanating from such a region would be modified to an extent depending on the local disturbance; and also that the reaction, to an external field, of an element of charge thus situated would be similarly affected. Under these circumstances the force between two positive electrons at a given distance apart will not necessarily be equal to the force between two negative electrons of equal charge separated by the same distance, on account of the difference of the masses; neither will the attraction between unlike electrons necessarily be equal to the repulsion between like electrons carrying equal charges.

Unfortunately there is an apparently fatal objection to any view which makes the modified law of force depend on what is practically nothing more than an alteration in the effective charge produced by the local energy or mass. For the law of force will then involve the product of two factors each of which depends on the state of only one of the two separate elements. If dm denotes the mass which is associated with the element of charge de, we can take $\frac{\partial m}{\partial e}$ as a measure of the intensity of the local disturbance. Using Taylor's Theorem we see that the force between two elements ∂E and $\partial e'$ instead of being equal to $\frac{\partial E \partial e'}{4\pi r^2}$ will be

$$\frac{\partial E \partial e'}{4\pi r^2}\left\{1 + a\frac{\partial m'}{\partial e'} + A\frac{\partial M}{\partial E} + aA\frac{\partial m'}{\partial e'}\frac{\partial M}{\partial E} + \dots\right\}\dots\dots\dots(4),$$

where A and a are fundamental constants which must be equal, by symmetry, when the charges are of like sign; otherwise this is not necessarily the case. Keeping the capital letters for positively charged elements and the small letters for negative charges it follows that, to the accuracy covered by the terms above, the force between mixed elements $(\partial E + \partial e)$ and $(\partial E' + \partial e')$ is

$$\frac{1}{4\pi r^2}\{\partial E + \partial e + A\,\partial M + a\partial m\}\{\partial E' + \partial e' + A\,\partial M' + a\partial m'\} \quad \ldots(5).$$

Assuming that the condition for neutrality is

$$\partial E' + \partial e' = \partial E + \partial e = 0,$$

the force between neutral systems becomes

$$\frac{1}{4\pi r^2}(A\,dM + a\,dm)(A\,dM' + a\,dm') \quad \ldots\ldots\ldots\ldots(6).$$

dM, dm, dM' and dm' are positive but otherwise arbitrary except that $\frac{dm}{dM} = \frac{dm'}{dM'}$, so that (6) is essentially positive. This corresponds to a repulsion and is incompatible with an attraction unless either A or a has an imaginary part.

If we are to get an attraction it is necessary that the change due to the masses should not enter as a separate factor for each of the elements of charge concerned, provided we are limited to functions whose expansions contain only real constants. The law of force may, for instance, be of the form

$$\frac{\partial e\partial e'}{4\pi r^2} f\left(\frac{\partial m}{\partial e}\,\frac{\partial m'}{\partial e'}\right) = \frac{\partial e\partial e'}{4\pi r^2}\left\{1 + a\frac{\partial m}{\partial e} + b\frac{\partial m'}{\partial e'} + c\left(\frac{\partial m}{\partial e}\right)^2 \right.$$

$$\left. + g\frac{\partial m}{\partial e}\frac{\partial m'}{\partial e'} + h\left(\frac{\partial m'}{\partial e'}\right)^2 + \ldots\right\} \quad \ldots\ldots(7),$$

using Stirling's Theorem. As before, $a=b$ and $c=h$ by symmetry, for charges of the same sign, but there is no longer any necessary relation between a and g. If the constants have the same value independently of the sign of the charges concerned, i.e. if

$$a = A = b = B \text{ and } c = C = h = H,$$

then the expression above leads to the following value for the force between two mixed systems $(\partial E + \partial e)$ and $(\partial E' + \partial e')$:

$$\frac{1}{4\pi r^2}\left\{(\partial E + \partial e)(\partial E' + \partial e') + a\left[(\partial E + \partial e)(\partial M' + \partial m')\right.\right.$$

$$\left. + (\partial E' + \partial e')(\partial M + \partial m)\right] + c\left[(\partial E + \partial e)\left(\frac{\partial M'}{\partial E'}\partial M' + \frac{\partial m'}{\partial e'}\partial m'\right)\right.$$

$$\left. + (\partial E' + \partial e')\left(\frac{\partial M}{\partial E}\partial M + \frac{\partial m}{\partial e}\partial m\right)\right]$$

$$\left. + g(\partial M + \partial m)(\partial M' + \partial m')\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8).$$

If the condition for neutrality is $\partial E + \partial e = \partial E' + \partial e' = 0$, the force between neutral systems is

$$g\,(\partial M + \partial m)\,(\partial M' + \partial m')/4\pi r^2 \quad \ldots\ldots\ldots\ldots\ldots(9).$$

As g may be negative, this agrees with the Newtonian law.

According to formula (8) the terms involving a and c might make an important difference to the force exerted on an electron by a gravitational field, but would have no effect on the attraction of two neutral particles, which is determined entirely by the term containing g. The introduction of this term is practically equivalent to attributing the Newtonian law of mass attraction to each pair of elements of charge however small.

Conditions for Neutrality.

The introduction of gravitational forces complicates electrostatic problems very considerably. The condition for electrical equilibrium in any conductor is that the stream of positive electricity in any direction at any point should be equal to the stream of negative electricity in the same direction at that point. In general this condition may be very complicated. If we may suppose that only the negative electrons are free to move, as the positive current is zero the negative current must be zero also. If we are also free to neglect effects arising from differences of pressure of the electrons at different points, the condition will be satisfied if at every point the electric force on an electron is exactly balanced by the gravitational force, the resultant average force on an electron then being zero. There will thus be a potential gradient in a conductor in equilibrium in a gravitational field; this potential gradient is, however, too small to be detected experimentally.

The way to define a neutral material particle is not an obvious matter under these circumstances. We might, for example, define a neutral material particle as one which exerted no force on a negative electron; but equally valid reasons could be urged for making the vanishing of the force on a positive electron the condition for neutrality. If both these conditions were satisfied there would be no force between two neutral particles, so that they are not compatible with the existence of gravitational attraction. In accordance with the considerations of the last section we might define a neutral particle as one which would give rise to forces exerting equal accelerations on both positive and negative electrons. Taking as an illustration the formulation of the forces on p. 592 the force on a negative electron at B, due to a particle consisting of N_1 positive and n_1 negative electrons at A, may be written

$$\frac{e}{4\pi r^2}\{N_1 E(1+c)+n_1 e(1+b)\}$$

and the acceleration will be obtained on dividing this expression by m. In a similar manner the acceleration of a positive electron at B, due to the material particle at A, is

$$\frac{E}{M \,.\, 4\pi r^2}\{N_1 E(1+a)+n_1 e(1+c)\} \quad \text{.........(10)}.$$

If these two accelerations are equal and if $E=-e$

$$\frac{N_1}{n_1}=\frac{bM-cm}{cM-am} \quad \text{........................(11)}.$$

Under these suppositions N_1/n_1 must have this ratio for the particle at A to be neutral. Previously we supposed N_1 and n_1 were equal. If we still suppose this to be the case we see that $c=(bM+am)/(M+m)$. Since M exceeds m, this is compatible with $2c>a+b$ if a exceeds b.

By attributing gravitational attraction to a slight excess of the attraction of the oppositely charged elements over the repulsion of the elements with like charges we are led to the paradoxical result that the addition of sufficiently small amounts of electrification of like sign to each of two neutral particles will increase the attraction between them instead of diminishing it. Consider, as an example, the specification of the forces which we have just been discussing. The repulsive force between two neutral particles is

$$\frac{E^2}{4\pi r^2}[N_1 N_2(1+a)+n_1 n_2(1+b)-(N_1 n_2+N_2 n_1)(1+c)].$$

If a small negative charge is given to both the particles n_1 will become $n_1+\delta n_1$ and n_2 will become $n_2+\delta n_2$, N_1 and N_2 being unaltered. Taking $N_1=n_1$ and $N_2=n_2$ as the condition for neutrality, the additional repulsion is

$$\{\delta n_1 \delta n_2(1+b)+(n_1\delta n_2+n_2\delta n_1)(b-c)\}\, E^2/4\pi r^2$$

or to the first order of small quantities

$$(n_1\delta n_2+n_2\delta n_1)(b-c)\, E^2/4\pi r^2 \quad \text{...............(12)}.$$

If equal positive charges were added, instead of negative charges, the additional repulsion would be

$$(n_1 \delta n_2 + n_2 \delta n_1)(a - c) E^2/4\pi r^2.$$

Since $2c$ exceeds $a + b$, either $a - c$ or $b - c$ must be negative. Thus at least one of the cases considered gives rise to an attraction. This attraction will only be experienced when very small quantities of electricity are added, since unity is very large compared with a, b or c. Thus the neglected second order term, which corresponds to the usual electrostatic repulsion, will very soon overcome the attractive forces under discussion.

If the negative electrons are alone capable of motion, the interior of a large solid conductor in equilibrium under its own gravitational field will contain a slight excess of negative electricity, since the gravitational force on an electron has to be supported by that arising from the electric intensity. If we neglect the pressure of the electrons, this condition will hold right up to the surface whatever the total charge on the conductor may be. Any electric or magnetic effects arising from this excess of negative charge would be very minute, even with conductors of the magnitude of the earth.

General Considerations.

A review of the preceding discussion shows that the electron theory is not in a position to make very definite assertions about the nature of gravitational attraction. It seems likely that the Newtonian law of attraction between elements of matter is one between elements of mass or confined energy and that it is of a very fundamental character. It is doubtful if it can be replaced by a modified law of electrostatic force between electrons or elements of electric charge, unless the modified law includes the associated mass explicitly. Even so, the case does not appear very simple. A number of alternative possibilities could be eliminated if the acceleration of a negative electron in the earth's field of gravitational force could be determined experimentally.

If gravitational attraction is, as it were, an uncompensated residue of the electrical forces, we should expect it, like all electrical actions, to be propagated with the velocity of light. Lorentz*, who has considered this question very fully, finds that if gravitation is propagated in the same manner as electrical actions it will give rise to effects practically identical with those which follow from the usual Newtonian law. The differences which arise are too small to be detected from astronomical data and they are also incapable of accounting for the recognized irregularities in the motions of the heavenly bodies.

---

* *Proceedings Akad. van Wet. te Amst.* vol. II. p. 559 (1900).

The view that gravitational attraction is an electrical effect at bottom is quite old. The definite form of it, that particles of uncharged matter contained equal and opposite charges and that the attraction between the unlike charges slightly exceeded the repulsion between the like charges, seems to have been first put forward by Mossotti*. The application to electrons in atoms has been considered by J. J. Thomson†.

The Relative Theory of Gravitational Effects.

The following speculations, due to Einstein‡, about the relation between gravitation and some other phenomena, are of considerable interest. The discussion sets out from the empirical law that in a uniform field of gravitation all material bodies move with equal accelerations. We have seen that this law is likely to be of a very fundamental character. Rather similar results have been obtained by Abraham§ by a somewhat different argument.

Let us consider two separate regions of space. In the first is a uniform field of gravitational force. This space is provided with a set of axes κ, at rest, and the lines of gravitational force run in the negative direction parallel to the z axis. The magnitude of the gravitational acceleration is γ. The second region of space is free from gravitational attraction and is provided with a set of axes κ' which move with uniform acceleration γ in the positive direction along the z' axis. The equations of motion of any particle in either system (provided the action of other material particles on it may be disregarded) are then the same, if the equations for each system are referred to the set of axes which we have associated with that system. They are clearly

---

* **O. F. Mossotti,** ***Sur les forces qui régissent la constitution intime des corps.*** **Turin, 1836.**

† ***Camb. Phil. Proc.*** **vol. xv. p. 65 (1908).**

‡ ***Ann. der Physik,*** **vol. xxxv. p. 898 (1911).**

§ ***Physik. Zeits.*** **13 Jahrg. p. 1 (1912).**

$$\frac{\partial^2 x}{\partial t^2}=0,\quad \frac{\partial^2 y}{\partial t^2}=0,\quad \frac{\partial^2 z}{\partial t^2}=-\gamma.$$

Now let us assume that our two systems are physically identical, i.e. that it is impossible to distinguish between them by any physical means whatever. This is tantamount to denying absolute acceleration in the same way that the principle of relativity denies absolute velocity. The equivalence of the two systems is obvious if we confine ourselves to the Newtonian mechanics. This is not the case however when we turn to some of the results which we have deduced from the principle of relativity.

It is shown on p. 318 that the mass of any system involves a part E/c^2, where E is the radiant electromagnetic energy of the system and c is the velocity of light. E/c^2 is mass in the sense of a coefficient of inertia but we have not proved that it is subject to gravitational attraction, in the way that the ordinary mass of the system is. If this mass is not subject to gravitational attraction it is evident that the equivalence of the systems κ and κ' will not hold exactly; systems in a uniform gravitational field will only fall uniformly provided the inertia of their electromagnetic energy may be neglected. A formal proof that if κ and κ' are exactly equivalent it is necessary that E/c^2 should represent gravitating mass, as well as inertia, is easily constructed

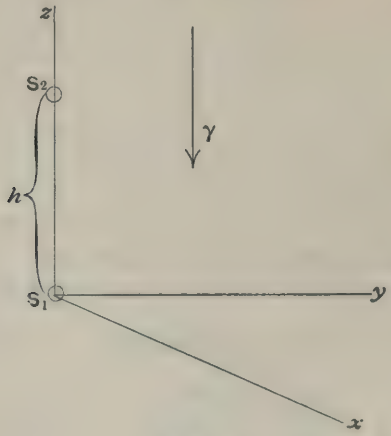


Fig. 58.

Let the diagram represent the axes κ in the field of uniform gravitational force. Let S_2, S_1 represent two sufficiently minute material systems at a distance h apart along the z axis. If S_2 and S_1 are small enough, h will remain invariable. The systems will simply slide along the z axis with uniform acceleration, as though they were held apart by a rigid connection of length h. Suppose that S_2 is allowed to send a certain amount of radiant energy E which is received by S_1. The energy is measured at S_2 and S_1 by instruments which can be brought together and compared. As we do not yet know anything about the effect of gravitation on electromagnetic energy we are unable to say anything about what takes place during the transference. We can, however, find out a great deal about it if we admit that the system κ' is equivalent to the system κ. For in the system κ' let us measure the energy of the radiation with reference to a set of axes κ_0 that are at rest with reference to κ' at the instant at which the radiation leaves S_2. If the energy thus measured is E when it leaves S_2, it will, by the theory of relativity*, be equal to $E(1+\gamma h/c^2)$ by the time it is at S_1. For S_1 is then moving with reference to the axes κ_0 with the relative velocity $v=\gamma t=\gamma h/c$. This result is true only to the first order of the small quantity $\gamma h/c^2$.

On the hypothesis that the two systems are equivalent, exactly the same result will have to hold with reference to the system of axes κ in the gravitating system. In this case γh is equal to ϕ, the difference in the gravitational potential between the points S_2 and S_1; so that if E_2 and E_1 are the energies of a given quantity of radiation at S_2 and S_1 respectively, then

$$E_1=E_2(1+\phi/c^2).$$

If we are to retain the principle of the conservation of energy in these cases it follows that work equal to $E_2\phi/c^2$ is done on the radiant energy by the gravitational forces. This is the same as if the radiant energy possessed gravitational mass equal to E/c^2. We therefore conclude that E/c^2, which the principle of relativity shows to be the inert mass of radiation, represents its gravitating mass also.

---

* Einstein, *Ann. der Physik* vol. XVII. p. 913 (1905).

The Velocity of Light in the Gravitational Field.

Returning to the system κ', let ν_2 be the frequency of the light emitted by S_2; then by Doppler's principle the light received by S_1 will appear to have a different frequency, viz. $\nu_1 = \nu_2(1 + \gamma h/c^2)$ to the first approximation. If the systems κ and κ' are equivalent, the gravitational field will therefore produce a change in the frequency ν_2 of the radiation which leaves S_2 to the value

$$\nu_1 = \nu_2(1 + \phi/c^2)$$

at a place S_1 where the gravitational potential differs by ϕ from its value at S_2. These results are true when the clocks used to measure time at S_2 and S_1 agree with one another when brought together. But the distance h is constant in the κ system; so that, if the radiation were being continually emitted by S_1 and received by S_2, there would be a continuous accumulation or depletion of waves in the space between them, if ν_1 and ν_2 were unequal. This contradiction indicates that the times at different parts of a gravitational field are not correctly given by clocks which agree when brought together. To give times which do not lead to a contradiction the clock at S_2 must go $1 + \phi/c^2$ times slower than the clock at S_1 when they are brought together and compared. Measured with such clocks the frequencies at S_2 and S_1 become the same and the number of waves in the stretch h is independent of the lapse of time, when the emission and reception are steady and continuous.

Now if we measure the velocity of light in different places of the system κ' with clocks which agree when compared together we get always the constant quantity c. If κ and κ' are equivalent, the same is true of the gravitating system κ. But these clocks do not go at the right rate in different parts of κ; whence it follows that the velocity of light is not the same at different parts of κ but varies with the gravitational potential according to the relation

$$c_1 = c_2\left(1 + \frac{\phi}{c_1^2}\right)$$

Thus the principle of the constancy of the velocity of light which Einstein made the basis of the principle of relativity does not hold for gravitational fields according to this theory.

It is easy to calculate the curvature of a ray of light produced by gravitation on this view. Take two points on the wave front at distance dl apart. At time dt later the wave front will be tangential to spheres of radii $c\,dt$ and $\left(c+\frac{\partial c}{\partial l}\,dl\right)dt$ described about the two points. The rotation $\delta\theta$ of the wave front in time δt is thus

$$\delta\theta=\frac{\partial c}{\partial l}\,\delta t.$$

If δn is an element of the normal to the wave front $c\,dt=\delta n$,

$$\frac{\partial\theta}{\partial n}=\frac{1}{c}\frac{\partial c}{\partial l}=\frac{1}{c^2}\frac{\partial\phi}{\partial l}$$

to a sufficient approximation.

Let us apply this to the case of light passing within a distance δ of the centre of a star of mass m. Let k be the gravitational constant, r the instantaneous distance of the ray of light from the centre of the star and ψ the angle between r and the perpendicular from the star on the straight path of the undeflected light. Then $\phi=km/r$ and the small total angular deflexion $\delta\theta$ is

$$\delta\theta=\frac{1}{c^2}\int\frac{\partial\phi}{\partial l}\,dn=-\frac{km}{c^2}\int_{-\pi/2}^{\pi/2}\frac{1}{r}\frac{\partial r}{\partial l}\frac{d\psi}{\cos\psi}$$

$$=-\frac{km}{c^2\delta}\int_{-\pi/2}^{\pi/2}\cos\psi\,d\psi=-\frac{2km}{c^2\delta}.$$

The deflexion is directed towards the star, so that if the light comes from a second star the apparent angular distance between their centres will be increased; in the case of the light from a star passing just outside the sun's disc it amounts to 0·83 second of arc.

In a later paper Einstein* has extended the investigation so as to cover the relation between gravitation and electromagnetic phenomena other than mass and the velocity of radiation, using similar methods. The smallness of the expected effects appears to preclude the possibility of their experimental verification in the near future.

* *Ann. der Physik*, vol. XXXVIII. p. 355 (1912).

INDEX OF NAMES

Abraham, M. 122, 228, 230, 231, 239, 242, 262, 282, 483, 601
Adams, E. P. 439
Airy, Sir G. B. 134, 271, 275, 285
Ampère, A. M. 86, 362
Arago, F. 271 ff., 275, 283

Back, E. 523
Baedeker, K. 439, 440, 453, 457, 463, 466
Baeyer, O. v. 468
Balmer, J. J. 513, 514, 528 ff., 585, 586
Barkla, C. G. 487, 488, 490, 497, 499, 501, 502, 505, 509
Bartoli, A. 334
Beatty, R. T. 504
Beck, P. N. 397
Becquerel, J. 539
Berg, O. 452
Bergmann, A. 517
Berzelius, J. J. 1
Bessel, F. W. 33
Bevan, P. V. 527, 528, 533
Blondlot, R. 122
Bohr, N. 395, 413, 434, 530, 535, 585, 587, 594
Boltzmann, L. 334, 352, 379, 399, 508, 551
Boscovich, R. G. 270
Bradley, J. 270
Bragg, W. H. 477 ff., 481, 490, 498, 509
—— W. L. 508, 509
Broglie, P. De 509
Brown, F. C. 6, 442
Bucherer, A. H. 239, 242, 316
Buisson, H. 551
Byerly, W. E. 33

Campbell, N. R. 273
Cauchy, A. L. 155
Cavendish, Hon. H. 39, 43
Coblentz, W. W. 159
Cole, A. D. 124
Compton, K. T. 474
Cooke, H. L. 441
Corbino, O. M. 539
Cornu, M. A. 123
Coulomb, C. A. 12, 42, 43
Crowther, J. A. 493, 495, 496
Curie, J. 548
—— P. 375, 378, 390, 392, 394, 548

Darwin, C. G. 509
Davy, Sir H. 1
De Broglie, M. 509
Debye, P. 360, 507
Deslandres, H. 517
Diesselhorst, H. 411
Doppler, C. 526, 551, 604
Drecq, M. 355
Drude, P. 134, 136, 137, 407, 432, 438, 539, 547
Duane, W. 122
Du Bois, H. 392
Dufour, A. 524
Dulong, P. L. 357

Einstein, A. 295 ff., 314, 322, 354, 357, 359, 360, 424, 473, 474, 601, 603 ff.
Ettingshausen, A. von 434

Fabry, C. 122, 551
Fajans, K. 583
Falk, K. G. 579
Faraday, M. 1, 3, 39, 43, 86, 100, 121, 542, 545
Fermat, P. de 271
Féry, C. 355
Fitzgerald, G. F. 232, 280, 282, 283, 289, 291, 323
Fizeau, H. L. 273, 274, 276, 285, 306
Fleming, J. A. 124
Fourier, J. Baron 33, 327, 433
Fowler, A. 586
Franck, J. 532
Franz, R. 411, 423
Fredenhagen, K., 442
Fresnel, A. 271, 273 ff., 285
Friedrich, W. 507, 508

Garbasso, A. 535

Gauss, K. F. 20 ff., 28, 30, 41, 42, 43, 98, 204
Gehrcke, E. 535
Gehrts, A. 469
Geiger, H. 5, 494, 583
Gerlach, W. 355
Gray, J. A. 485
Green, G. 23, 32, 117, 190
Gutton, C. 122

Hagen, E. 139, 432
Hall, E. H. 409, 434 ff., 463
Hallo, J. J. 539
Hallwachs, W. 469
Havelock, T. H. 161
Heaviside, O. 13, 15, 218
Helmholtz, H. v. 2
Hertz, H. 122, 469
Herweg, J. 509
Hevesy, G. v. 583
Hicks, W. M. 515
Himstedt, F. 122
Hollnagel, H. 159, 359
Honda, K. 392
Horton, F. 467
Hughes, A. Ll. 475
Hull, G. F. 215
Humphreys, W. J. 533
Hurmuzescu, D. 122

Jaeger, G. 411
Jamin, J. C. 134
Jeans, J. H. 38, 80, 103, 105, 122, 343, 345, 399, 411, 422, 423, 430, 434, 550 ff., 554
Jentzsch, F. 441
Joule, J. P. 106, 429, 433

Kamerlingh Onnes, H. 378, 424, Preface
Kaufmann, W. 7, 235, 239, 316
Kaye, G. W. C. 485
Kayser, H. 513, 517
Keesom, W. H. Preface
Kelvin, Lord 425, 430, 459, 460, 550, 570, 571, 581, 585
Kerr, J. 545, 547
Kirchhoff, G. 189, 331, 469, 512
Kirschbaum, H. 534
Klatt, V. 531
Kleeman, R. D. 478
Knipping, P. 507
Koenigsberger, J. 462, 467, 468
Kundt, A. 154
Kurlbaum, F. 355
Küstner, H. 469

Ladenburg, R. 526, 527, 528
Lagrange, J. L. 105, 175
Lamb, H. 231, 275
Langevin, P. 251, 259, 362, 374, 379, 380, 392
Laplace, P. S. Marquis de 18, 26, 30, 32, 33, 42, 43
Larmor, Sir J. 2, 109, 170, 229, 258, 286, 291, 522, 533, 549
Laue, M. 325, 507
Lebedew, P. 215
Le Bel, J. A. 546
Lecher, E. 452
Leduc, A. 434
Lees, C. H. 412
Lenard, P. 482, 525, 531
Leroux, F. P. 154
Liebreich, E. 441
Liénard, A. 244
Lindemann, F. A. 358, 360
Livens, G. H. 533
Lodge, Sir O. J. 276
Lommel, E. 530
Lorentz, H. A. 2, 3, 149, 156, 161, 162, 167, 177, 183, 228, 233, 239, 241, 276, 282, 286, 289 ff., 295, 315, 316, 345, 373, 374, 413, 433, 522, 524, 525, 535, 600
Lorenz, L. 149, 411, 423
Loria, S. 527, 528
Lummer, O. 356

McLaren, S. B. 434
Maclaurin, R. C. 162
MacLean, G. V. 122
Maculoso, D. 539
Marsden, E. 494
Maurain, C. 393
Maxwell, J. C. 2, 6, 14, 35, 81, 96 ff., 101, 102, 107, 115, 121, 122, 142, 149, 155, 207 ff., 212, 215, 223, 224, 230, 232, 271, 307, 413 ff., 417, 508, 551
Merritt, E. 530
Michelson, A. A. 274, 277, 278, 280, 282
Millikan, R. A. 5, 6, 356
Minkowski, H. 325
Mohler, J. F. 533
Morley, E. W. 274, 277, 278, 280, 282
Moseley, H. G. J. 509, 510, 511
Mossotti, F. O. 601

Nagaoka, H. 550
Nelson, J. M. 579
Nernst, W. 358, 360, 434, 463
Newton, Sir I. 7
Nichols, E. F. 158, 159, 179, 180, 215
—— E. L. 530
Nicholson, J. W. 395
Nicol, J. 501

Oersted, H. C. 86
Ohm, G. S. 97, 106
Onnes, H. Kamerlingh 378, 424, Preface
Oxley, A. E. 392

Paschen, F. 517, 522, 523
Pasteur, L. 546
Patterson, J. 436, 440
Pellat, H. 122

Peltier, J. C. 412, 426, 429, 430, 436, 455 ff., 459, 462, 463
Perot, A. 122
Perrier, A. 378
Perrin, J. 6
Petit, A. T. 357
Pickering, E. C. 586
Planck, M. 267, 276, 322, 344, 347 ff., 352, 354 ff., 360, 395, 407, 424, 473 ff., 585
Poisson, S. D. 26, 28, 30, 31, 42, 43, 47, 52, 82, 119, 218
Poynting, J. H. 203 ff., 208, 213, 215, 221, 484, 520
Preston, T. 522
Priestley, J. 12
Pringsheim, E. 356
Purvis, J. E. 522

Rankine, A. O. 282
Rayleigh, Lord 149, 155, 175, 343, 345, 347, 350, 355, 412, 434, 550
Riecke, E. 407
Righi, A. 434
Ritz, W. 517, 529, 530
Roentgen, W. C. 477, 478, 482
Rosa, E. B. 122
Rossi, R. 584
Routh, E. J. 464, 492
Rowland, H. A. 99
Rubens, H. 139, 158, 159, 359, 432
Runge, C. D. T. 517, 522
Russell, A. S. 583, 584
Rutherford, Sir E. 5, 6, 356, 482, 490, 494 ff., 583, 584, 594
Rydberg, J. R. 511, 517, 529, 551, 586

Sadler, C. A. 497, 499, 504
Saeland, S. 531
Saunders, C. A. 122
Schmidt, G. C. 532
Schott, G. A. 550, 552, 554
Schoute, C. 452
Schuster, A. 432, 483
Schwarzschild, K. 535
Sellmeier, W. 149, 155, 156, 161, 177, 179, 181
Soddy, F. 583, 594
Sommerfeld, A. 325, 483, 484
Stark, J. 8, 513, 517, 526, 531, 534
Starke, H. 7
Stefan, J. 333, 334, 341, 355
Steinberg, K. 439
Steubing, W. 531
Stewart, B. 331, 469, 512
Stirling, J. 353, 402, 597
Stokes, Sir G. G. 4, 91, 95, 101, 271, 275 ff., 280, 283, 483, 502, 530, 532
Strutt, Hon. R. J. 594
Stuhlmann, O. 478, 479, 481

Thomson, Sir J. J. 3, 4, 7, 122, 208, 209, 218, 229, 232, 234, 242, 249, 407, 422, 423, 426, 429, 430, 433, 438, 440, 456, 458, 482, 483, 485, 490, 491, 494 ff., 505, 529, 550, 554, 560, 570, 579 ff., 584, 585, 601
Townsend, J. S. 4
Trouton, F. T. 282
Trowbridge, J. 122

Voigt, W. 524, 535, 539, 544, 545, 547
von Baeyer, O. 468
von Ettingshausen, A. A. 434

Walden, P. 575
Weber, W. 362, 407
Webster, A. G. 44, 223
Wehnelt, A. 441
Weiss, P. 380, 381, 384, 385, 387, 388, 390, 394, 395, 397, 423, 462
Wendt, G. 534
Whiddington, R. 504, 505, 509, 586
Whittaker, E. T. 12, 276, 283, 530
Wiechert, E. 3, 234, 244, 483
Wiedemann, E. 532
—— G. 411, 423
Wiedmann, G. 469
Wien, W. 327, 346, 354, Preface
Williamson, B. 317
Wilson, C. T. R. 3, 4
—— H. A. 5, 277, 283, 430, 431, 434, 442, 529
Wood, R. W. 528, 532, 533, 535, 539

Zeeman, P. 2, 3, 373, 375, 513, 518, 521 ff., 527, 528, 533 ff., 539, 542, 545, 547

INDEX OF SUBJECTS

α rays, scattering of 490
Aberration of light 269, 306
Absorbing media 163
Absorption, feeble 166; of characteristic X rays 497 ff.
Absorption of light 144; coefficient of 165; near critical frequencies 178
Acceleration and force 260
Aether 268; and the principle of relativity 323
Aether pulse theory of the Roentgen rays 483
Airy's experiment 271
Alloys, conductivity of 465
Arago's experiment 271
Atom, definiteness of 549; of Kelvin and Thomson 554
Atomic, numbers 510, 583; structure 491, 494, 549, 554, 564; theory, application to the 6; weights of elements, and gravitational attraction 592; weights of radioactive elements 583
Atoms, Bohr's theory of 395, 585; number of electrons in 488, 494, 495, 564
Attraction, gravitational, compared with electrostatic 590
Avogadro's number 6
Axes, moving 286, 297

β rays, deflexion by electric field 236; deflexion by magnetic field 237, 240; scattering of 490
β secondary X rays 503
Band spectra 517
Black body radiation 330, 433
Bohr's theory of atoms 395, 585; theory of spectra 585
Bucherer's experiment 239

Characteristic X rays, absorption of 497 ff.
Charge, accelerated, field due to 247, 249; accelerated, radiation from 257; central, of atom 494, 510, 583, 585; electric, with variable velocity 243; on electron, determination of 3, 356
Charged system, in uniform motion 217
Charges, fictitious 57
Chemical combination 570; and transference of electrons 571, 574, 580
Collisions between particles under central forces 417
Combination, chemical 570
Conditions at boundary between two dielectric media 45
Conducting electrons 463
Conducting media, propagation of electromagnetic waves in 135; reflexion at boundary of 137
Conduction in metals, electron theory of 408, 413, 463; simple theory of 409
Conductivity, electric, of alloys 465; electric, of bad conductors 467; electrical 410, 420, 422; electrical, for periodic forces 430; of metals for heat 411, 421
Constant, Boltzmann's 406; Planck's 356, 469, 506, 509, 585
Contact difference of potential 454
Contractile electron 232
Contraction, Fitzgerald's 280, 291
Correlation, principle of 286, 290
Corresponding states, of magnetization 382
Coulomb's Law 23
Critical temperature, magnetic 389
Crystals and X rays 507
Curie's Law 378
Current, convection 98; displacement 97; electric 96; force on an element of, in magnetic field 114; magnetic force due to element of electric 113; polarization 98
Currents, induction of 100

Diamagnetism 361, 368
Dielectric constant, electron theory of 74
Dielectric media 39; Poisson's theory of 47; electron theory of 58
Differential equations satisfied by vectors when charges are present 186

Dispersion, theory of 142, 152; anomalous 154; of rock-salt 161; generalized theory of 169; of luminous gases 526
Displaced electrons, potential due to 67
Displacement Law, Wien's 343
Distribution of energy, molecular 403
Distribution of potential energy 404
Doppler effect 284, 304
Double refraction, magnetic 539, 544
Doublet, potential due to 58, 69

Electricity, different elements of 9; specific heat of 429, 448, 451, 452
Electrified particles, force between 10
Electrodynamics, first law of 95; second law of 101
Electromagnetic, equations 307; inertia 7, 228 ff., 234
Electromagnetism 86; dynamical theory of 102
Electron, accelerated 199; at rest 197; contractile 232; determination of charge on 3, 356; expanding 552; field due to rigid 222; in uniform motion 197, 218; positive 8, 482, 580, 593; relativity theory of mechanics of 310; revolving, magnetic force due to 362, 367; theory, origin of 1
Electronic, conduction, kinetic theory of 398; conductors, equilibrium theory of 441, 444
Electrons, conducting 463; emission of, from hot bodies 441; in equilibrium with conductors, temperature relation 444, 447, 450, 454; law of force between 591; number of, in atoms 488, 494, 495, 564; reflexion of, by conductors 468
Electro-optics 534, 547
Energy, electric, of charge in uniform motion 220, 252; in electric field 34, 44; inertia of 316; in field of accelerated point charge 251, 255; magnetic, of charge in uniform motion 221, 252; of a system of charges 33; of moving particle 312; of oscillator 349; rate of loss of, by aether 203
Entropy 399; and probability 400; of resonators 253
Equations, electromagnetic, for moving systems 286, 290, 307; Laplace's 18, 42; Lorentz's 183; Poisson's 26, 42; of line of force 16; of propagation 115; universal 182

Ferromagnetic substances, specific heats of 397
Ferromagnetism 361, 380, 393, 397
Field due to uniformly moving charges 218
Fitzgerald's contraction 280, 291
Fizeau's experiment 273, 306
Fluorescence 530
Force, exerted on electric charges 205; thermoelectromotive 459
Forces, activity of the 201; between material particles 595, 599; continuously operative, on principle of relativity 320

Galvanomagnetic effects 434
Gas, paramagnetic 379
Gases, luminous, dispersion of 526
Gauss's theorem 20 ff., 41
Gravitation 588, 600; effect of, on light 601, 604; relative theory of 601
Green's theorem 23

Hall effect 434, 437

Induction, electric 40, 51; magnetic 82
Inertia of energy 316
Intensity, electric 14, 20; of reflected light, minimum value of 181
Inverse Zeeman effect 535
Iron, magnetic properties of 391
Isotropic radiation, pressure exerted by 211

Jeans and Rayleigh's radiation formula 343

Kaufmann's experiment 235
Kerr's electro-optical effect 547; magneto-optical effect 545
Kinetic theory and thermodynamics 399
Kirchhoff's Law 330; solution 189

Laplace's equation 18, 42; operator, transformation of 26
Law, Curie's 378; of Stewart and Kirchhoff 330; of Wiedemann-Franz-Lorenz 411
Least time, principle of 271
Light, aberration of 269, 306; velocity of, in the gravitational field 604
Line of force, equation of 16

Magnetic, double refraction 539, 544; field, effect of, on electric currents 434, 437, 439; force due to moving electron 362; properties, abrupt changes in 390; rotation of plane of polarization 535, 542
Magnetism 77; application of thermodynamics to 377; on electron theory 361
Magnetization, mechanical reaction caused by 395; permanent 381, 387
Magneto-optics 373, 518, 523, 535, 539, 545
Magneton 394

Mass, electromagnetic, of rigid electron 228; electromagnetic, variation with velocity 234; longitudinal and transverse 229 ff.; variation with velocity on the principle of relativity 312
Material particles, conditions for neutrality 598; forces between 595, 599
Maxwell's stresses 35, 81
Mechanics of electron 310
Michelson and Morley's experiment 277
Moment, magnetic, of revolving electron 367
Momentum, and radiation pressure 212; effect of inertia of energy on 320; electromagnetic 208, of charge in uniform motion 221, of isolated system 215, of rigid electron 227; of moving particle 314, 319
Moving axes 286, 297

Natural rotation of plane of polarization 545
Negative electron 7
Neutral material particle 598
Newton's law of gravitation 588
Number of electrons in atoms 488, 494, 495, 564
Numbers, atomic 510, 583

Optical, convection, Stokes's theory of 275; effects due to motion of refracting medium 273
Oscillator, energy of 349

Paramagnetic gas 379
Paramagnetism 361, 376, 379
Peltier effect 429, 455, 457
Periodic law, and radioactive elements 583; and structure of atoms 565
Permeability, diamagnetic 371
Phosphorescence 530
Photoelectric action 469
Planck's radiation formula 347, 354, 355
Point charge in uniform motion, field due to 248
Poisson's equation 26, 42; solution of equations of propagation 118; theory of dielectric media 47
Polarization, electric 51; electron theory of 64, 70; magnetic rotation of plane of 535, 542; natural rotation of plane of 545; of X rays 488; potential due to 60; variable 54
Positive electricity, structure of 580
Positive electron 8, 482, 580, 593
Potential, at internal points 19; contact difference of 454; electric 15; magnetic, due to electric current 86; propagated 193; scalar 194; vector 194, 197
Poynting's theorem 203
Pressure of radiation 209
Pressure shift of spectral lines 533
Principle of correlation 286, 290
Principle of least time 271
Principle of relativity 296, 322
Probability and entropy 400
Probability, of distribution of energy among oscillators 352, 357; of statistical distribution of a gas 401
Propagation, equations of 115
Pulse theory of Roentgen rays 483
Pyrrhotite, magnetic properties of 384

Quasi-stationary motion 262

Radiant energy, gravitational mass of 603
Radiation, and temperature 326, 433; complete 328, 433; energy density of complete, in refracting medium 332; formula of Planck 347, 355; formula of Rayleigh and Jeans 343; formula of Wien 354; from accelerated electric charge 257; isotropic 211; pressure and momentum 212; pressure of 209; reaction of, on accelerated charge 263; reflexion of, at a moving mirror 277, 335; secondary, asymmetrical emission of 477 ff.; types of 476
Radioactive elements and periodic law 583
Ratio of electromagnetic to electrostatic unit 122
Rayleigh and Jeans's radiation formula 343
Rays, residual 157
Reaction of radiation on moving charge 263
Reflecting power, of metals 140; of quartz 159
Reflexion, change of phase on 138, 168; of electromagnetic waves 129; of electrons by conductors 468; of light by moving mirror 277, 335; of light near critical frequencies 178
Refracting medium, optical effects due to motion of 273; theory of propagation of light in moving 284
Refraction, of compounds and mixtures 150; of electromagnetic waves 129
Refractive index 149, 165
Relative theory of gravitation 601
Relativity, principle of 296; dependence of physical quantities on velocity 322; experimental test of 242, 315, 318
Residual rays from absorbing media 167
Resistance, change in a magnetic field 439
Reststrahlen 159
Roentgen rays 482 ff., 496 ff., 507 ff.

Scattering, of α and β rays 490; of α and β rays, Rutherford's theory 494; of α and β rays, Thomson's theory 491; of X rays 485

Secondary rays 477, 496
Series in band spectra 517
Series, spectral 513, 585
Shell, force on a magnetic 83
Shells, polarized 62
Specific heat of electricity 429, 448, 451, 452
Specific heats, at low temperatures 358; Debye's theory 360; Einstein's theory 357; of ferromagnetic substances 397
Spectra, band 517
Spectra, Bohr's theory of 585; X ray of elements 510
Spectral emission, theories of 525, 585
Spectral lines, definiteness of frequency of 551; displacement in electric field 534; effect of pressure on 533
Spectral series 513, 585
Spectroscopic phenomena 512, 517, 525, 551, 585
Stability, of electrons in atom 552, 554, 558, 562
Stark effect 534
Stefan's law 333
Stirling's approximation 402
Stokes's theorem 91; theory of optical convection 275
Stresses, Maxwell's 35, 81, 207 ff.
Structural formulae, in chemistry 575
Structure of positive electricity 580

Temperature and radiation 326, 433
Thermionics 441
Thermodynamics and kinetic theory 399
Thermoelectricity 425, 448, 455, 459, 461
Thermoelectromotive force 459
Thermomagnetic effects 434
Thomson effect 429, 448, 452
Transparency of quartz 159
Tubes of force 16, 22

Unit of current 89
Unit of electric charge 13
Units, electrical 109

Valency, chemical, and atomic structure 566, 573
Values, actual and mean 65
Velocity, addition of, on principle of relativity 302; of light in the gravitational field 604; of propagation of electro-magnetic field 121; of secondary β rays emitted by characteristic X rays 504; wave 252

Wave of reorganization 258
Wave, plane-polarized electromagnetic 124
Wave-length of X rays 510
Waves, electromagnetic 115
Wien's law 337; displacement law 343; radiation formula 354

X ray spectra 510
X rays, and crystals 507; β secondary 503; characteristic, absorption of 497 ff.; characteristic, velocity of secondary β rays emitted by 504; polarization of 488; scattering of 485; wave-length of 510

Zeeman effect 373, 518, 523; inverse 535; theory of 373, 518

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS

CAMBRIDGE PHYSICAL SERIES

The Times.—"The Cambridge Physical Series...has the merit of being written by scholars who have taken the trouble to acquaint themselves with modern needs as well as with modern theories."

Modern Electrical Theory. By N. R. Campbell, Sc.D. Second edition, largely rewritten. Demy 8vo. pp. xii + 400. 9*s*. net.

"The treatment throughout is admirably clear and readable, the arrangement is logical, and the work of different investigators is carefully considered and given due weight....Possibly the most striking feature is the completeness of the book, for it would be hard to find a section of the subject which has not received due attention."—*Cambridge Review* (on the Second Edition)

"Mr Campbell's work should be read by all students of physical science; those who desire merely a good general knowledge of recent developments in electricity will find that it contains all that is essential; while those who wish to specialise will find it an excellent introduction and guide."—*Electrical Engineering* (on the First Edition)

Sound. An Elementary Text-Book for Schools and Colleges. By J. W. Capstick, M.A., D.Sc., Fellow of Trinity College, Cambridge. Crown 8vo. pp. viii + 296. With 120 figures. 4*s*. 6*d*.

"An important addition to the *Cambridge Physical Series*....The latter portion is of especial value, and gives an interesting *rationale* of the principal orchestral instruments....Dr Capstick has succeeded in making a remarkably lucid exposition of his theme, and the book will, we have little doubt, promptly take its place as a standard text-book on the subject."—*Guardian*

Mechanics. By John Cox, M.A., F.R.S.C. Demy 8vo. pp. xiv + 332. With four plates and 148 figures. 9*s*. net.

"Prof. Cox has his eye specially upon mechanical principles, avoiding merely mathematical difficulties so far as that is fairly possible; he 'starts from real problems, as the subject started, showing how the great investigators attacked these problems, and introducing the leading concepts only as they arise necessarily and naturally in the course of solving them'; he 'brings out incidentally the points of philosophic interest and the methods of science'; he appeals constantly to experiment for verification, 'leading up to an experimental course limited to the most important practical applications,' and eventually embodying a good deal of matter not usually found in the elementary text-books; and he adds limited sets of carefully selected examples for exercise. Students that will not learn mechanics from this work, and be fired with interest in the subject, must be hopeless. Students that must work through an ordinary text-book for examinational or other reasons ought to have this volume by their side for its interest and stimulus, and for its fruitful co-ordination of theory and practice."—*Educational Times*

2000
7.15

CAMBRIDGE PHYSICAL SERIES

The Study of Chemical Composition. An Account of its Method and Historical Development, with illustrative quotations. By Ida Freund. Demy 8vo. pp. xvi + 650. With 99 figures. 18*s*. net.

"The accomplished Staff Lecturer on Chemistry of Newnham College has admirably succeeded in her efforts to produce a book dealing with the historical development of theories regarding chemical composition....No part of the book is dull, and the student who starts on it will be led on to continue to the end, to his own great advantage. Many people have no time or opportunity to refer to a large mass of original papers, and the value of this book of careful selections carefully and cleverly put together will be very great....Students of physical science who wish to form a clear conception of the laws of chemical combination, of the meaning and accuracy of combining weights and combining volumes, of the doctrine of valency and of isomerism, and of the parts played by various investigators and authors in elucidating these and kindred matters cannot do better than refer to this book."—*Athenaeum*

Mechanics and Hydrostatics. An Elementary Text-book, Theoretical and Practical, for Colleges and Schools. By R. T. Glazebrook, C.B., M.A., F.R.S., Director of the National Physical Laboratory and Fellow of Trinity College, Cambridge. Crown 8vo. 6*s*.

Also in separate volumes

Part I. **Dynamics.** pp. xii + 256. With 99 figures. 3*s*.
Part II. **Statics.** pp. viii + 182. With 139 figures. 2*s*.
Part III. **Hydrostatics.** pp. x + 216. With 98 figures. 2*s*.

"A very good book, which combines the theoretical and practical treatment of mechanics very happily....The discussion of force, momentum, and motion is consistent and clear; and the experiments described are rational and inexpensive. This treatment is calculated to give much clearer ideas on dynamical concepts than a purely mathematical course could possibly do."—*Journal of Education on* Dynamics

"A clearly-printed and well-arranged text-book of hydrostatics for colleges and schools....The descriptions are clearly written, and the exercises are numerous. Moreover, the treatment is experimental; so that altogether the book is calculated to give a good grasp of the fundamental principles of hydrostatics."—*Nature on* Hydrostatics

Heat and Light. An Elementary Text-book, Theoretical and Practical, for Colleges and Schools. By R. T. Glazebrook, C.B., M.A., F.R.S. Crown 8vo. 5*s*.

Also in separate volumes:

Heat. pp. xii + 230. With 88 figures. 3*s*.
Light. pp. viii + 210 + vi. With 134 figures. 3*s*.

"The very able author of the treatise now before us is exceptionally well qualified to deal with the subject of theoretical and experimental physics, and we may at once say that he has succeeded in producing a class-book which deserves, and will doubtless receive, a full share of the patronage of our school and college authorities."—*Mechanical World*

CAMBRIDGE PHYSICAL SERIES

Electricity and Magnetism: an Elementary Text-book, Theoretical and Practical. By R. T. GLAZEBROOK, C.B., M.A., F.R.S. Crown 8vo. pp. vi + 440. 6*s.*

"As an elementary treatise on the laws of electricity and magnetism it leaves little to be desired. The explanations are clear, and the choice of experiments, intended to carry home these explanations to the mind of the student, is admirable. We have no doubt that teachers of the subject will find the volume of great use. It is the work of one who has had long experience as a teacher in an important centre of education and all who study its contents carefully and intelligently will acquire a sound knowledge of the principles of the attractive subject with which it deals."—*Engineering*

Photo-Electricity. By A. LL. HUGHES, D.Sc., B.A., Assistant Professor of Physics in the Rice Institute, Houston, Texas. Demy 8vo. pp. viii + 144. With 40 text-figures. 6*s.* net.

"The book goes very thoroughly into all modern experimental research related to its subject....The whole question is important and may have far-reaching results, and we consider the author is to be heartily congratulated on the lucid manner in which he has dealt with it."—*Electrical Review*

The Electron Theory of Matter. By O. W. RICHARDSON, D.Sc., F.R.S., Wheatstone Professor of Physics at King's College, London. Demy 8vo. pp. viii + 612. With 58 figures. 18*s.* net.

"This book will serve as an admirably compact and yet very complete account of the subject, leading directly to many points where research, theoretical and experimental, is proceeding."—*Science Progress*

A Treatise on the Theory of Alternating Currents. By ALEXANDER RUSSELL, M.A., M.I.E.E., Principal of Faraday House. In two volumes. Demy 8vo. Vol. I. pp. xiv + 534. Second edition. 15*s.* net. Vol. II. pp. xii + 488. 12*s.* net.

"It reveals the thorough mastery which the author has of the subject, and the capable reader will find much that is of value. Recent contributions are carefully digested and generally elucidated, and, on the whole we may consider the book an up-to-date treatment of the alternating-current theory."—*Electrical Review*

Experimental Elasticity. A Manual for the Laboratory. By G. F. C. SEARLE, Sc.D., F.R.S., University Lecturer in Experimental Physics. Demy 8vo. pp. xvi + 187. 5*s.* net.

"The author...gives us in this book a collection of the experiments which he has found suitable for University students in their second and third years' preparation for the Natural Sciences Tripos....The book can be strongly recommended to all teachers and to University students. It forms an excellent monograph on elementary elasticity."—*Guardian*

CAMBRIDGE PHYSICAL SERIES

Air Currents and the Laws of Ventilation. Lectures on the Physics of the Ventilation of Buildings. By W. N. SHAW, Sc.D., F.R.S., Fellow of Emmanuel College, Director of the Meteorological Office. Demy 8vo. pp. xii + 94. 3*s.* net.

"The present volume by Dr Shaw is most welcome, because it draws attention forcibly to the physics of the ventilated space, and deals with the whole subject in a thoroughly scientific manner....The treatment of the subject is quite original....The whole volume is highly suggestive, and constitutes a real advance in the study of this important subject."—*Engineering*

Conduction of Electricity through Gases. By Sir J. J. THOMSON, O.M., D.Sc., LL.D., Ph.D., F.R.S., Fellow of Trinity College, Cambridge, and Cavendish Professor of Experimental Physics. Second edition enlarged and partly re-written. Demy 8vo. pp. viii + 678. 16*s.*

"It is difficult to think of a single branch of the physical sciences in which these advances are not of fundamental importance. The physicist sees the relations between electricity and matter laid bare in a manner hardly hoped for hitherto....The workers in the field of science are to-day reaping an unparalleled harvest, and we may congratulate ourselves that in this field at least we more than hold our own among the nations of the world."—*Times* (on the First Edition)

A Treatise on the Theory of Solution, including the Phenomena of Electrolysis. By WILLIAM CECIL DAMPIER WHETHAM, Sc.D., F.R.S., Fellow of Trinity College, Cambridge. Demy 8vo. pp. x + 488. 10*s.* net. [*New edition in preparation*

"The present work embraces practically all the material on the subject of solutions which is dealt with in the ordinary text-books of physical chemistry, except that part concerned with velocity of reaction and purely chemical equilibrium. The treatment throughout is characterised by great clearness, especially in the physical and mathematical portions, so that the volume may be warmly recommended to students of chemistry who desire to increase their knowledge of this department of the subject."—*Nature*

The Theory of Experimental Electricity. By W. C. D. WHETHAM, Sc.D., F.R.S. Second edition. Demy 8vo. pp. xii + 340. 8*s.* net.

"This book is certain to be heartily welcomed by all those who are engaged in the teaching of theoretical electricity in our University Colleges. We have no hesitation in recommending the book to all teachers and students of electricity."—*Athenaeum* (on the First Edition)

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
London: FETTER LANE
Edinburgh: 100, PRINCES STREET
ALSO
London: H. K. LEWIS, 136, GOWER STREET, W.C.

PLEASE DO NOT REMOVE
CARDS OR SLIPS FROM THIS POCKET

UNIVERSITY OF TORONTO LIBRARY

P&A Sci.